

AD-A096 025 KANSAS UNIV/CENTER FOR RESEARCH INC LAWRENCE REMOTE --ETC F/G 17/9

RADAR IMAGE INTERPRETABILITY ANALYSIS.(U)

JAN 81 V S FROST, J A STILES, J C HOLTZMAN

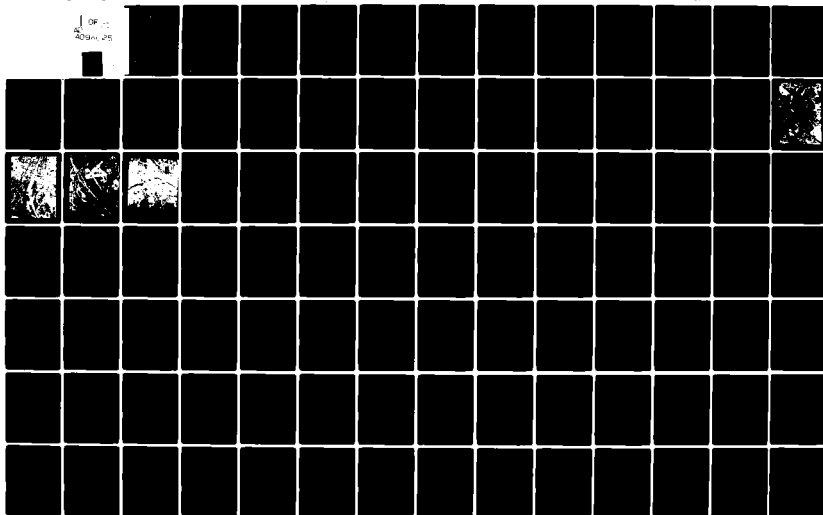
DAAG29-77-6-0075

UNCLASSIFIED CRINC/RSL-TR-342-2

ARO-14043.11-65

NL

1 OF 12
AD-A096 025



ARO 14043.11-AS
REMOTE SENSING LABORATORY

CRINC
|| ■ ||

LEVEL II

12

AD A 096025

DDC FILE COPY



THE UNIVERSITY OF KANSAS CENTER FOR RESEARCH, INC.

2291 Irving Hill Drive—Campus West
Lawrence, Kansas 66045

81 3 2 070



THE UNIVERSITY OF KANSAS CENTER FOR RESEARCH, INC.

2291 Irving Hill Drive—Campus West
Lawrence, Kansas 66045

Telephone: (913)864-4832

RADAR IMAGE INTERPRETABILITY ANALYSIS.

9 Final Report. 1 May 77 - 7 Apr 80

10

V. S. Frost, J. A. Stiles and J. C. Holtzman

11

January 1981

12 141

U. S. Army Research Office

15

Grant Number DAAG29-77-G-0075

14

CRIMP RSL-TR-342-2

Remote Sensing Laboratory
The University of Kansas
Center for Research, Inc.
Lawrence, Kansas 66045

13 HFO

14 14043.11-G

Approved for Public Release;
Distribution Unlimited

40668



REMOTE SENSING LABORATORY

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. <i>AD-A096 025</i>	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Radar Image Interpretability Analysis		5. TYPE OF REPORT & PERIOD COVERED Final Report 5/1/77-4/7/80
		6. PERFORMING ORG. REPORT NUMBER 342-2
7. AUTHOR(s) V. S. Frost, J. A. Stiles and J. C. Holtzman		8. CONTRACT OR GRANT NUMBER(s) DAAG29-77-G-0075
9. PERFORMING ORGANIZATION NAME AND ADDRESS Remote Sensing Laboratory University of Kansas Center for Research, Inc. Lawrence, Kansas 66045		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE January 1981
		13. NUMBER OF PAGES 101
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) NA		
18. SUPPLEMENTARY NOTES The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Radar Images, Image Quality, Synthetic Aperture Radar, Image Interpretation		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The utility of radar images with respect to trained image interpreter ability to identify, classify and detect specific terrain features (linear, natural area, complex area features, and individual man-made features) was qualitatively determined. Further, radar images were evaluated with respect to their utility for determining vehicle movement potential and the level of activity within the test areas. Because there are no physical laws governing the relationship between image quality and measurable image characteristics, signal-to-noise ratio, dynamic range, image bandwidth, geometric fidelity and root-mean-square error, a second		

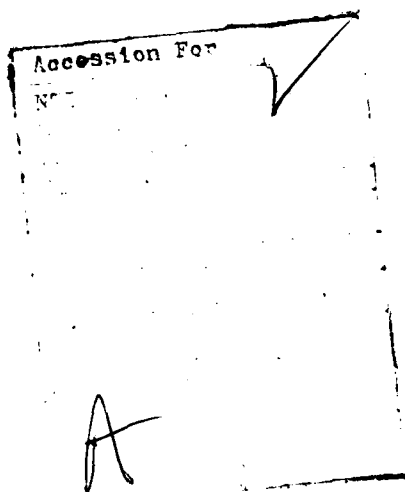
DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. order linear statistical model was assumed. An experiment was designed and conducted to estimate the coefficients of the model given each image application. The result of this analysis was that statistically significant models were obtained to relate the measured image properties to the interpreters' ability to analyze linear features and to evaluate the utility of radar images for vehicle movement potential and activity level. Further, it was found that the relative importance of the measured image properties with respect to image utility changed with image application. This study has provided useful information as to how certain image characteristics relate to radar image utility as a function of several image applications.



ABSTRACT

The utility of radar images with respect to trained image interpreter ability to identify, classify and detect specific terrain features (linear, natural area, complex area features, and individual man-made features) was qualitatively determined. Further, radar images were evaluated with respect to their utility for determining vehicle movement potential and the level of activity within the test areas. Because there are no physical laws governing the relationship between image quality and measurable image characteristics, signal-to-noise ratio, dynamic range, image bandwidth, geometric fidelity and root-mean-square error, a second order linear statistical model was assumed. An experiment was designed and conducted to estimate the coefficients of the model given each image application. The result of this analysis was that statistically significant models were obtained to relate the measured image properties to the interpreters' ability to analyze linear features and to evaluate the utility of radar images for vehicle movement potential and activity level. Further, it was found that the relative importance of the measured image properties with respect to image utility changed with image application. This study has provided useful information as to how certain image characteristics relate to radar image utility as a function of several image applications.

TABLE OF CONTENTS

	<u>Page</u>
1.0 INTRODUCTION	1
1.1 Background	3
1.1.1 The Notions of Quality and Interpretability for SAR	4
1.1.2 Previous Quality and/or Interpretability Work for Radar Imagery	6
1.1.3 Radar Speckle and Multiplicative Noise Literature	9
1.1.4 Interpretability/Quality Literature in the Optics Field	13
1.1.5 Data Base (Test Target) Design for Interpretability/Quality Studies	14
2.0 EXPERIMENTAL DESIGN	16
3.0 DEGRADATION PROCEDURES	23
4.0 RADAR IMAGE QUALITY METRICS	26
5.0 CORRELATION ANALYSIS	32
6.0 REGRESSION ANALYSIS	40
6.1 The Linear Model and Results	41
6.1.1 Mathematical Basis	41
6.1.2 Model Parameter Estimation	42
6.1.3 Analysis of the Prediction Equation	45
6.1.4 Blocking	49
6.1.5 Results of the Linear Regression Analysis	52
6.2 Non-Linear Models and Results	66
7.0 CONCLUSIONS	70
REFERENCES	73
APPENDICES: A Interpreter Packet	A-1
B Published Papers and Scientific Personnel	B-1

LIST OF FIGURES

	<u>Page</u>
Figure 1	18
Figure 2	19
Figure 3	20
Figure 4	21

LIST OF TABLES

	<u>Page</u>
Table 1: Correlation Coefficient for Each Interpreter	35
Table 2: Correlation of Each Interpreter With the "Average Interpreter"	36
Table 3: Correlation Between the Average Interpreter and the Degradation Parameters and Quality Metrics	38
Table 4: Evaluation of Predicted Regression Models	53
Table 5: Evaluation of the Importance of Image Quality Parameters	56
Table 6: Regression Coefficients for Response Category 1 for the Full Model	58
Table 7: Regression Coefficients for Response Category 1 for the Reduced Model (No Blocking)	59
Table 8: Regression Coefficients for Response Category 1 for the Reduced Model (No Blocking and No $x_i x_j$ terms)	60
Table 9: Regression Coefficients for Response Category 2 for the Reduced Model (No Blocking and No $x_i x_j$ or x_i^2 terms)	61
Table 10: Regression Coefficients for Response Category 5 for the Full Model	62
Table 11: Regression Coefficients for Response Category 5 Reduced Model (No Blocking)	63
Table 12: Regression Coefficients for Response Category 5 Reduced Model (No Blocking or $x_i x_j$ terms)	64
Table 13: Regression Coefficients for Response Category 5 Reduced Model (No Blocking, $x_i x_j$ or x_i^2 terms)	65
Table 14: Image Quality Based on the Spatial-Grey-Level Volume	69

1.0 INTRODUCTION

The human thought processes involved in the interpretation of an image field $I(x,y)$, which may have been generated by any one of several types of imaging devices (for example, an aerial photography camera, an imaging radar, an infrared scanner), are infinite in variety. The interpretation rules which each of us have stored mentally differ according to our past experiences and to our own distinct methods of dissecting and analyzing visual material. There is some consistency, however, among human judgments of image quality and interpretability (especially for photographs).

The goal of a quality measure is the evaluation of the behavior of an imaging system in as concise terms and in as few parameters as is possible. Quality metrics may be used by system designers or by the image user; the former is knowledgeable about the individual components of the sensor but may not think of the system output in terms of its images while the latter knows his application's needs, so he knows in image terms what he wants to get out of the system. The need for quality metrics which one can quickly reference increases greatly with increased complexity of the system (e.g., more lenses, mirrors, signal processing filters, and so forth).

The notion of an image quality measure is especially appealing for radar system designers because one finds that the thought processes involved in designing an imaging radar (real or synthetic aperture) can tend to exclude in-depth image application considerations. Worries about power consumption, source stability, antenna cross-polarization levels, etc., can rapidly overwhelm the design engineer leaving the "goodness" or utility of the ultimate product, the images to chance. Thus the

primarily mathematical nature of his task and the size of his task may preclude learning very much about radar imagery interpretation.

Radar image quality metrics which relate to his choice of design parameters may help fill a gap in the range of experience. Quality measures also can help image users who should be instrumental in mission planning.

Though there is a substantial amount of literature concerning photographic image quality measures (e.g., Pratt, 1978; Linfoot, 1960), only a few are applicable to the radar situation. Quite different kinds of information are conveyed by the photographs and the radar images as they employ disparate spectral bands. We specifically consider active microwave sensor images in this paper, for they are so distinct by nature of the process by which they are formed that quality measures derived for incoherent optical systems do not suffice. The basic discrepancies between the two image types arise from the varying degrees of illumination coherence and from geometrical and other spectral considerations. As an example of the causes of geometrical differences, the radar may consist of a monostatic arrangement while the illumination source and the sensor for the photograph are usually spatially separated. Additionally, the coherent imaging systems are often modeled as multiplicative noise processes, while incoherent image systems are often modeled with the noise being additive in effect. The speckle common to both coherent optical and radar systems dictates that the nature of quality measures and for example, image processing algorithms, should be different for radar images and ordinary incoherent optics photographs.

The development of relationships between the radar image metrics (e.g., dynamic range, signal-to-noise ratio, mean square bandwidth, etc.) and the image utility for several applications is a monumental task.

The experimental design which was adopted by the authors called for a number of trained radar/photo interpreters, numerous image metrics and image applications to achieve statistically significant results. In order to handle the volume of data collected and yet to maintain visibility of any group of variables of interest, response surface procedures were applied to the data analysis (Myers, 1971). This methodology allows one, for instance, to arrange in order of importance the image metrics given an application (assuming that the necessary supporting data were collected). Prior to presentation of these results we will discuss first the general principles of radar image formation and second, we will give a review of previous radar image quality studies.

It is hoped that the new results documented here will be not only interesting but also useful for those scientists and engineers whose tasks relate to imaging radar. The images of such devices are truly fascinating as they present a different way of "looking" at terrain, whether it is a portion of the surface of the Earth or of the surface of Venus. Understanding the information in the radar image and knowing how to design the active microwave system to gather relevant information are significant achievements; these were our goals at the onset of this study.

1.1 Background

The organization of the following sections is intended to stress image evaluation in the SAR context. Previous quality-interpretability work for SAR is discussed, along with the coherent speckle literature, and then work in the processing of images, containing signal dependent noise is covered. Analogous quality/interpretability work in the optical systems literature is reviewed. Test target considerations for image

evaluation are briefly described. But first it is worthwhile to talk about the concepts of image interpretability and quality for radar imagery.

1.1.1 The Notions of Quality and Interpretability for SAR

When one speaks of quality and interpretability investigations, it is obvious that even the definitions of these two terms are uncertain. Here several thoughts on the distinctions between the two, and also on their relations to one other, perhaps measurable, physical features or quantities are discussed. Pratt (1978) and Haralick (1978) have also considered the terminology difficulties implied here, in the context of digital image processing.

Foremost, we equate SAR image quality and interpretability to image fidelity and intelligibility, respectively, following Pratt's (1978) notation. The thoughts expressed below represent our opinions on the term quality (fidelity):

- (1) Image quality cannot be consistently, directly, equated to interpretability (intelligibility).
- (2) A high quality SAR system produces an output (an image) which selectively mimics the input (just prior to the antenna) in its spatial or spatial frequency structure; processing such as azimuth correlation, range compression, and speckle reduction are necessary in the production of an intelligible (to a human) scene.
- (3) Geometric fidelity is important for a quality image or imaging sensor (e.g., since processing of a slant range radar image is based partly upon the assumption of range perspective, then it is important for the image to conform to these expectations).

- (4) Quality measures do not rely on the previous existence of mental matched filter for objects in the scene (Barnard, 1972).
- (5) Quality does not particularly relate to size, shape, or orientation of features in the target scene.
- (6) Question - Is the quality relatable (analytically or empirically) to the complexity of the restoration filter which takes the scene back to the "ideal image"?
- (7) Question - How is quality (fidelity) related to the information content?
- (8) A fidelity measure can be applications independent, as opposed to an interpretability measure.
- (9) Existing quality measures can generally be broken into the univariate and bivariate types; the former consists of measurements made on a single image field while the latter involves numerical comparison between a pair of images (e.g., between the test and the "ideal" image).

Image interpretability or intelligibility can be understood to be distinct from quality (fidelity) in light of the following facts (and opinions):

- (1) Interpretability is related to the human observer's previous experience (with SAR imagery); the existence of a mental matched-filter is important (Barnard, 1972).
- (2) Geometric fidelity is less critical in general for intelligibility as opposed to fidelity measures (for modest geometric warpings).
- (3) Interpretability can be aided in some instances by stereoscopic or other special viewing capabilities.

- (4) Interpretability can be improved by appropriately designed enhancement or restoration filters.
- (5) Intelligibility can be improved by scale changes and scene rotations (Barnard, 1972).
- (6) Object recognition, related to interpretability, involves a time element (unlike quality).
- (7) Interpretability is related to the image coding scheme; for instance, negating the gray scale coding can impede recognition, compressing the dynamic range in the image can lead to poorer interpretability, etc.
- (8) Question - How is the level of interpretability related to the the complexity of a restoration filter?
- (9) Question - How are information content and image interpretability linked? A low information content scene can be highly interpretable and yet a great deal of information content does not guarantee a large "interpretability" factor.
- (10) Definitely, interpretability is applications-dependent.

These remarks have been introduced to preface the following review of literature on radar and optical quality and interpretability research. One notices that the interchanging use of these terms is widespread and little care is generally taken to define the concepts before pursuit of the various measures.

1.1.2 Previous Quality and/or Interpretability Work for Radar Imagery

The unclassified literature has been reviewed and a number of publications regarded to be significant will be discussed. In all cases but two, radar imagery examples did not accompany the published versions of

the quality and/or interpretability studies. Moore (1979) presented various examples of radar imagery, and R.L. Mitchell (1974) employed radar speckle simulations for his experiments.

W.A. Penn (1962) authored one of the earliest unclassified papers on signal fidelity for radar imagery. His discussions center upon "background" interference between output signal cells, multiplicative (fading) noise, and processing and display nonlinearities. Incoherent (post-detection) integration is introduced to lessen the fading variations. Through a heuristic argument, Penn suggests a measure of the radar map "quality" to be the product of $P \cdot Q$ where P is the number of samples averaged and Q is the average "signal-to-correlation noise." In Penn's notation, the correlation noise is larger than the amplitude of the noise by a factor of the time bandwidth product. Simulations employing aerial photographs were used for demonstration and for empirical derivation of the $P \cdot Q$ conclusion.

D.G. Corbett et al. (1964) developed a recognition metric for radar imagery target features. It was assumed that the inverse of the time taken (by equally trained observers) to reach a correct decision could be functionally related to size of the target and a number of transmissivity measures of the target and its background. A disappointingly low correlation coefficient was observed between the developed prediction equation and the target-recognition times when applied to new radar imagery (similar to that employed in the study).

R.O. Harger (1973) performed a theoretical study of SAR imagery that, although different from all the other interpretability studies mentioned herein, is included for completeness. While most SAR systems are designed around impulse response criteria, Harger suggests a design to minimize the probability of classification error (it is assumed that the field

to be classified is a known region whose boundaries are predetermined). Given the SAR system, noise and reflectivity density spectral density models assumed, the decision problem set up by Harger is relevant for a "Gaussian signal in Gaussian noise." The solution for the optimum classification role involves a nonlinear filter which includes a matched filter (i.e., the solution for impulse response design criteria).

R.L. Mitchell (1974) presents an interesting simulation experiment in which he begins with a cross on a dark background to study sensitivity of the SAR image with respect to resolution and averaging. The cross consists of many pixels whose individual grey shades are taken from sample functions of random distributions, e.g., Rayleigh, and log-normal for several different variances. Mitchell then varies the incoherent averaging, background noise, and film response. His conclusion is that image "quality" (which he is equating with the recognition of the cross shape) varies most directly with the image resolution. His results, though interesting and useful, would have been more realistic if the backgrounds had also been radar speckle patterns (from similar distributions, different means, etc.) rather than constant tone.

R.H. Mitchell (1974) developed a SAR image quality analysis model in his Ph.D. dissertation (University of Michigan.) His thinking is strongly influenced by the thorough review he presents on optical image quality measures. The "Radar Threshold Quality Factor" developed by Mitchell incorporates measures of the impulse response main lobe width, the signal-to-noise ratio, noncoherent integration, and human visual system factors. The RTQF is assumed to have a Gaussian shaped impulse response (effects of impulse response sidelobes are not included). Additionally, the clutter model employed is additive rather than multiplicative for the sake of mathematical tractability. Validation of the trends

predicted by the RTQF was accomplished by experimentation with a radar holographic viewer built by the author.

G.R. DiCaprio and J. Wasielewski (1976) performed an interpreter study using conventionally processed and incoherently degraded SAR imagery. Of interest was the target detection accuracy for the test scenes and the times required to perform the identifications. The conclusions of the authors were that the incoherent degradation allowed the radar interpreters to detect targets significantly faster and, secondly, that if more supporting ground truth data had been available the noncoherent degradation would have improved detection accuracy also.

D.W. Craig and M.L. Hershberger (1977) also reported results of an interpreter study. Employing trained observers, they investigated the effects of radar sensor, display, and mission variables on tactical target acquisition performance. Serious criticism of their results arises because target types, area coverage and display quality were dissimilar for the high/low resolution imagery cases. However, simply restating their conclusions, for their tasks and experimental set-up, higher resolution imagery performed superiorly to lower resolution imagery in all tests. R.K. Moore (1979) performed a sensitivity study to determine the effects on radar image interpretation of spatial resolution, using non-square pixels, and noncoherent averaging. This research will be discussed later in this document.

1.1.3 Radar Speckle and Multiplicative Noise Literature

In attempts to analytically or empirically deal with SAR image quality, interpretability, classification, edge detection, and so on, many investigators have performed research on the effects of the coherent nature of SAR imagery. The importance of speckle modeling has become apparent as techniques of wide band optical image processing (which do

not have to deal with small signal-to-noise ratios) have failed to be applicable to SAR imagery (e.g., especially edge detection algorithms). It seems particularly relevant to review the literature in this field in light of the fact that SAR impulse response quality criteria are deterministic in nature and yet random SAR speckle is recognized as a primary degradation factor in many interpretability studies.

J.W. Goodman (1976) presents some of the fundamental characteristics of speckle and derives the exponential probability density function for certain speckle situations. He also demonstrates the important fact that addition of M uncorrelated speckle patterns on an intensity basis improves viewing and reduces the image contrast (σ/μ) by $1/\sqrt{M}$. Time, space, frequency, or polarization diversity is utilized to obtain the independent speckle patterns, as is well known in SAR theory and operation. A good bibliography of historically interesting and contemporary articles concludes this paper.

A. Kozma and C.R. Christensen (1976) illustrate an experiment in which both a grating and a continuous tone image were illuminated coherently and incoherently. Subjective analyses showed that the speckle masks spatial information present in the image and has the effect of increasing the minimum resolution patch that can be obtained with a given aperture area. For imaging the grating (bar targets) it was determined subjectively - that the aperture of the coherent system needed to be about twice as large as that for an equivalent incoherent illumination system to obtain equal resolution. For the continuous tone target employed, the aperture ratios increased to approximately five. This difference between the two target types was hypothesized to be related to differences in complexity of the decision process. Noncoherent addition of independent speckle patterns to achieve a signal-to-noise ratio of about ten allowed

the coherent and incoherent system performances to be approximately equal as concerned apparent resolution capability. A similar result was found for the continuous tone target ($S/N \approx 10$). This agrees well with the findings of Moore (1979).

J.S. Zelenka (1976) discusses the mathematics supporting a discrete and a scanning system for SAR speckle reduction. As in many other discussions, the averaging is accomplished by processing subapertures coherently and then adding the resulting images on an intensity basis. The two methods described differ in that the discrete method forms non-overlapping independent apertures while the samples are not in this sense independent for the scanning case, though effectively there are more samples. The conclusion of Zelenka is that the scanning processor is more effective for speckle reduction in terms of signal-to-noise improvement vs. loss in resolution given a certain resolution of the final, output image.

Porcello et al. (1976) also treat the topic of image speckle reduction for SAR. Frequency and angular diversity are suggested in the mixed integration processor context. A series of noncoherently averaged radar images is presented to demonstrate improved viewing capabilities. The same basic information presented by Zelenka (1976) is given in this paper.

In addition to the studies of the statistical nature of coherent speckle, several articles have been published on image processing in the context of multiplicative noise. This research is reviewed because speckle is a form of multiplicative noise. Whereas the authors referenced thus far in this section have attempted speckle reduction at the point in radar image formation when amplitude and phase are both control-

table, the next few authors' (Frost et al., 1981) works examine the case of speckle reduction when only magnitude (no phase) data is available to the investigator; that is, image data (not signal film or radar holograms) was experimented upon.

Walkup and Choens (1974) and Kondo et al. (1977) discuss image processing and restoration by the use of a Wiener filter when the noise process is modeled as signal-dependent (rather than being treated as additive). The signal and noise are both considered to be wide sense stationary random processes, with the noise spectral density assumed known. The Wiener filter derived is non-adaptive (because of the stationarity assumptions). Experimental results indicate that after Wiener filtering is done to estimate the signal, edge detection algorithms applied on the output have a greater probability of representing true boundaries, rather than false boundaries characteristically produced on speckled, non-smoothed radar or photographic imagery. The above work was done for restoration of images corrupted by film grain noise.

Naderi and Sawchuk (1978) report the results of running adaptive Wiener filters on images degraded by film grain noise. The adaptive nature of the filter is necessitated because of the nonstationarity of the image first order statistics. The results presented demonstrate the improvements brought about by making the Wiener filters adaptive.

Oppenheim et al. (1968) discuss nonlinear filtering of multiplied and convolved signals to develop the "homomorphic" filter, not constrained by linearity assumptions, to produce an estimate of the desired signal. The homomorphic, Wiener, inverse, and constrained least squares estimation filters are discussed in a tutorial digital image processing article by B.R. Hunt (1975). The interesting result of his restoration

techniques shows the constrained least squares (CLS) filter to produce the most visually pleasing image from a degraded version of the same scene. Though theoretically the Wiener filter is the optimal linear filter, Hunt hypothesizes that the human visual system transfer function "matches" better with the output of the CLS and, thus, gives the human observer the impression of a superiorly reconstructed scene.

Summarizing this section, one finds that even for the basically "error-free" SAR system, in which fading noise is the dominant degradation factor, extracting information from or digitally restoring the imagery can be greatly aided by a priori knowledge of the fading statistics. Formation of noise models is an obviously important step in the design of the restoration (signal estimation) filter. Not only should the knowledge of speckle characteristics be applied when one develops SAR interpretability measures, but also, techniques for speckle noise removal will have application to correction of other types of SAR "errors". The formulation of an image restoration technique must also enter into the quantification of interpretability/quality measures of a SAR system degraded intentionally in the proposed study.

1.1.4 Interpretability/Quality Literature in the Optics Field

The understanding of imaging optical and radar systems is greatly enhanced by formation of a linear systems model (to a lesser extent by a nonlinear mathematical model). Both types of systems (radar and optical) have been characterized by a two-dimensional point spread function (psf); attempts at quantifying system quality by singling out measures of the psf have been less than successful (Brock, 1967).

The difficulties in using impulse response criteria or modulation transfer function criteria have been discussed in the optical systems

context by Brock (1967) and Noffsinger (1970,1971). Boiling down their discussions to the essence, one finds that quality/interpretability measures that do not factor in the application or the observation circumstances fail to adequately represent the system in question. Granger and Cupery (1972) attempt to incorporate the human factor in their article "An Optical Merit Function (SQF), Which Correlates with Subjective Image Judgments." A visual system model is used to develop the SQF. Similar visual system models are applied by Stockham (1972) in an excellent article and in Hunt (1975). Many other references, too numerous to mention, relevant to image quality/interpretability exist; a good bibliography is presented in Brock (1967). Mitchel (1974) selected several of these articles and others that he felt were pertinent to SAR quality.

1.1.5 Data Base (Test Target) Design for Interpretability/Quality Studies

Just as one would not test a recipe using poor quality ingredients, one must also carefully choose test targets for system analysis. A common mistake to be avoided in SAR image quality work is use of a test scene which is unsuitable because of its spectral composition. For example, many workers in the optical field have employed a tri-bar, or multiple bar target mistakenly believing that this target models a sine wave transmittance in the space domain and, thus, believing that its spectrum can be modeled as a delta function in the spatial frequency domain (Brock, 1967). Thus, for either simple or complex scenes the power spectra should be known, and should be suitable for the experiment at hand. For example, sufficient bandwidth, or "whiteness" of the spectrum of a target scenario might be important. If the spectrum does not satisfy one's criteria, it

can be augmented and the new scene defined by the inverse transform of the augmented spectral version.

T.W. Barnard (1972) discusses another aspect of target selection in his article "Image Evaluation by Means of Target Recognition." The emphasis is not on spectral content in this case, but rather he is concerned with providing targets to the human observers for which they already possess mentally stored visual "matched filters." Barnard gives the examples of the "Landolt-C," numerals, and "Spokes" targets. Similar in nature are the vision-testing charts containing letters of our alphabet, and those having the "E's" opening up, down, right, left. Barnard would e.g., define as unacceptable a chart of Cyrillic alphabet characters for the English speaking observers. This would seem to be an important consideration for SAR system interpretability if only humans not familiar with radar interpretation are available to rank or describe simulated scenes. Pratt (1978) also discusses target scene selection.

2.0 EXPERIMENTAL DESIGN

An experimental design in general consists of selecting the different conditions under which observations will be obtained. Proper selection of these conditions, i.e., fully exploring the likely range of operational conditions, is essential for efficient estimation of a relationship between the experimental variables and the response. In this case the experimental conditions refer to a specific level of image "quality". Therefore, to obtain radar images with controlled levels of utility a radar image was processed (degraded) using digital techniques. The response in this case is the judgment of human interpreters of the utility of the degraded radar images.

An experiment was designed to investigate the relationship between measured image quality parameters (IQP) and image utility for several applications. Because of the large number of data points (degraded images) required to totally explore desired relationships it was decided to assume that all third and higher order interactions between the IQP's could be deleted. That is, a second order model was assumed. This assumption has been successful in the past (Soliday, 1974; Williges, 1971; Craig and Hershberger, 1977).

Techniques outlined in Myers (1971) were used to efficiently specify the number and level (of degradation) of the required images. A central composite design (CCD) with uniform precision was selected. For example, with five image metrics, 32 different experimental conditions are required. The advantage of this approach was that the minimum number of observations were used and that each degradation was selected so that all the data points had uniform importance. Eight extra degraded images were added to comprise the data set.

Optically processed synthetic aperture radar imagery of the Roanoke, Virginia area was obtained in the form of film positive transparencies. The radar system was the Goodyear XA2 which has a resolution (6 dB width of the image impulse response) of approximately 15 ft. The transmitted carrier frequency was approximately 10 GHz and the system transmitted and received horizontal polarizations. The parameters of the flight which collected this imagery include a 20,000 foot altitude with a near range distance (ground track to near range) of 8 miles and a far range distance of 10 miles. The flight date was June 20, 1968. The scale of the received imagery was 1:100,000. The selected imagery was digitized using a sampling rate of 1000 pt/in.; this yields one pixel every 8.3 ft. which is close to the required Nyquist spacing of 7.5 feet. Any aliasing effects were thus ignored as were the effects of the sampling aperture. Four subareas of this imagery were selected for analysis because of the variety of targets contained therein. These scenes are shown in Figures 1-4; Figures 1 and 2 are aerial photographs and Figures 3 and 4 are the original radar images. The 40 test images were generated from these images. That is, a complete data set consisted of 10 degradations of scene A, 8 degradations of scene B, 10 degradations of scene C, and 12 degradations of scene D as prescribed by the CCD.

Each of eight trained Army photo-interpreters (PI) was given a complete set of 40 degraded images, a brief introduction to the experiment explaining its goals and the interpreter's role. Also included were instructions, questions, and interpretation guidelines. An answer sheet was provided for each image. The answer sheet allowed each PI to rank each image according to his ability to identify, classify, and detect specific terrain features. For analysis the PI's responses were averaged to form four basic response categories: (1) linear features, (2) natural area-

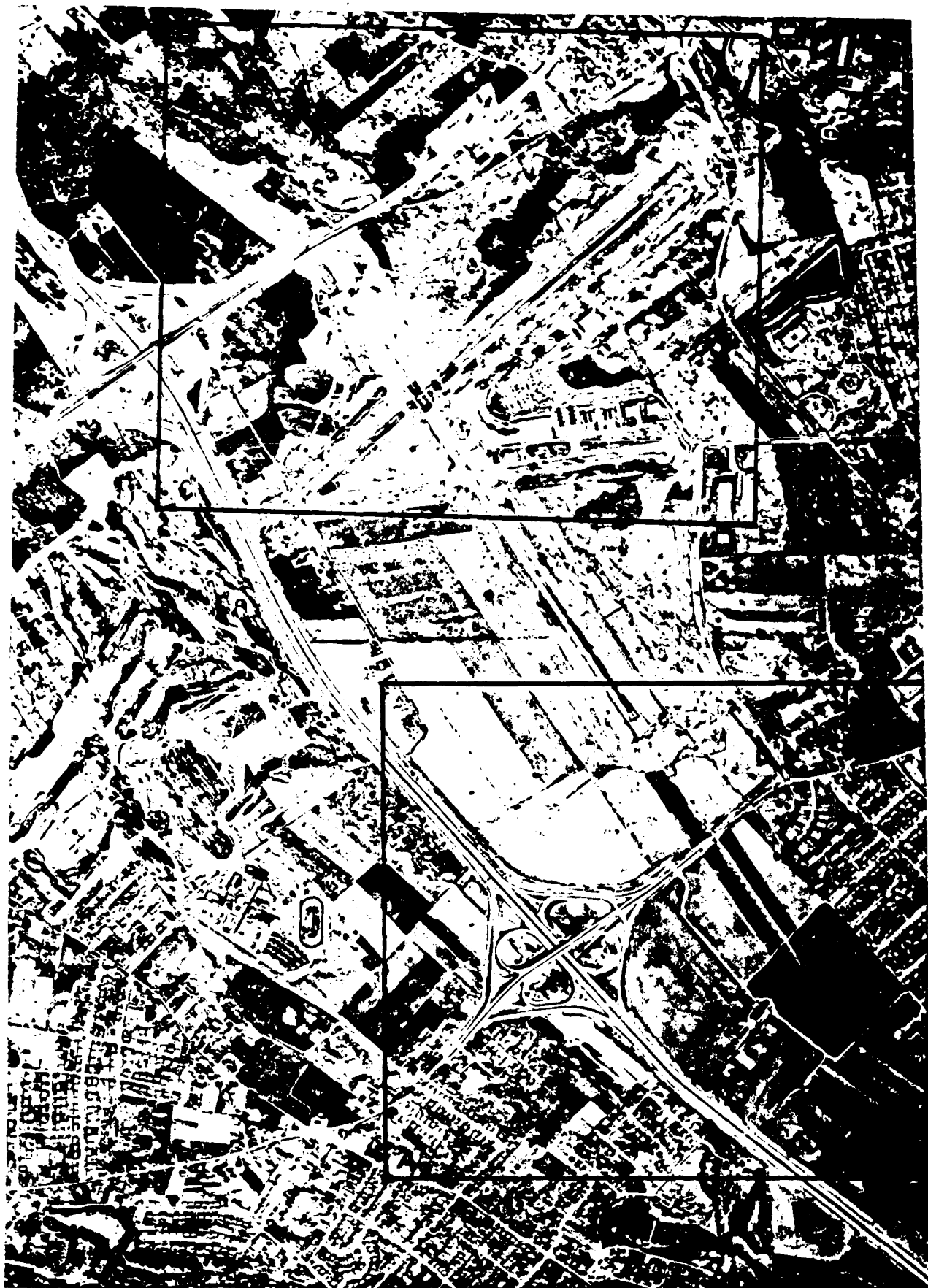


Figure 1



Figure 2

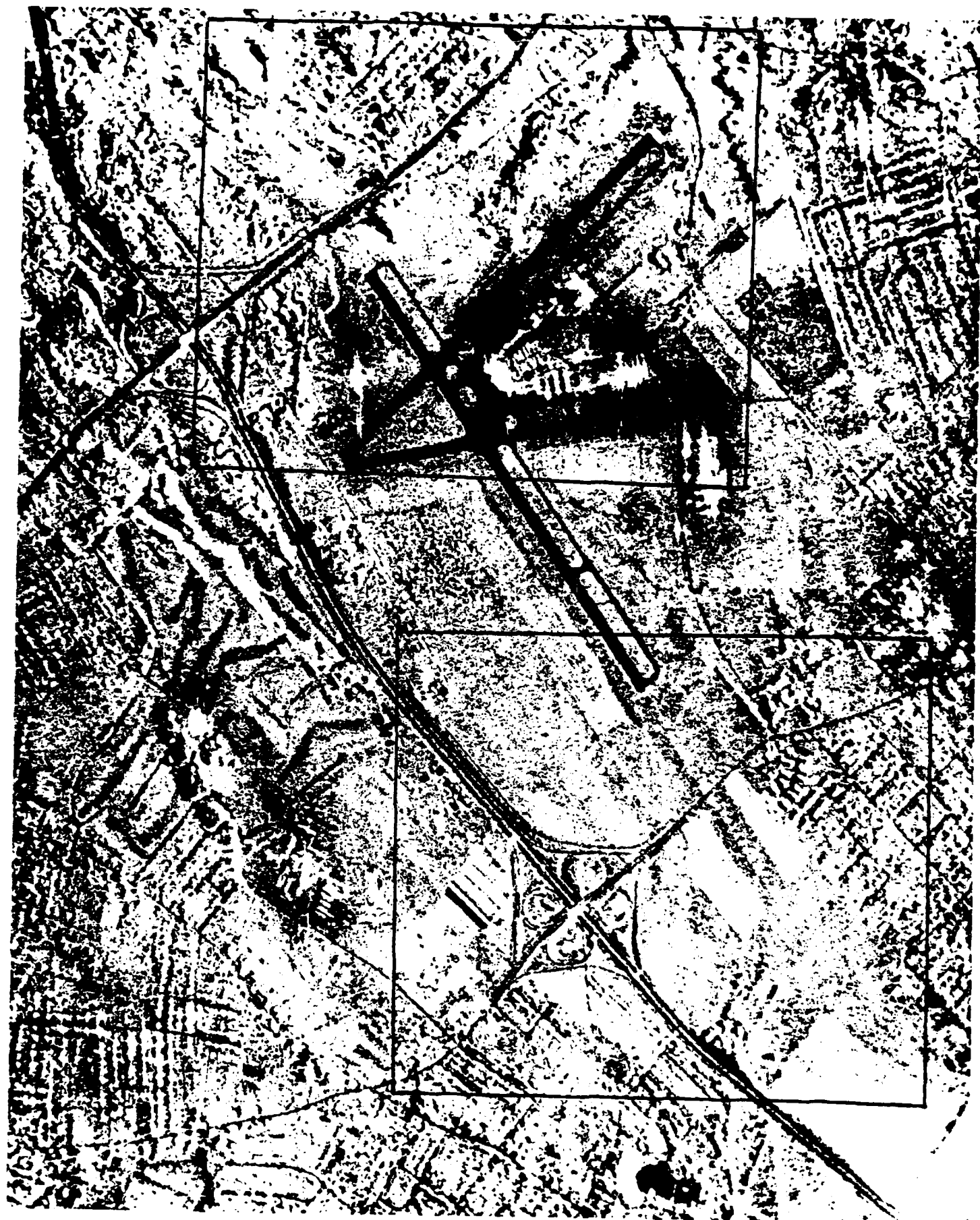


Figure 3

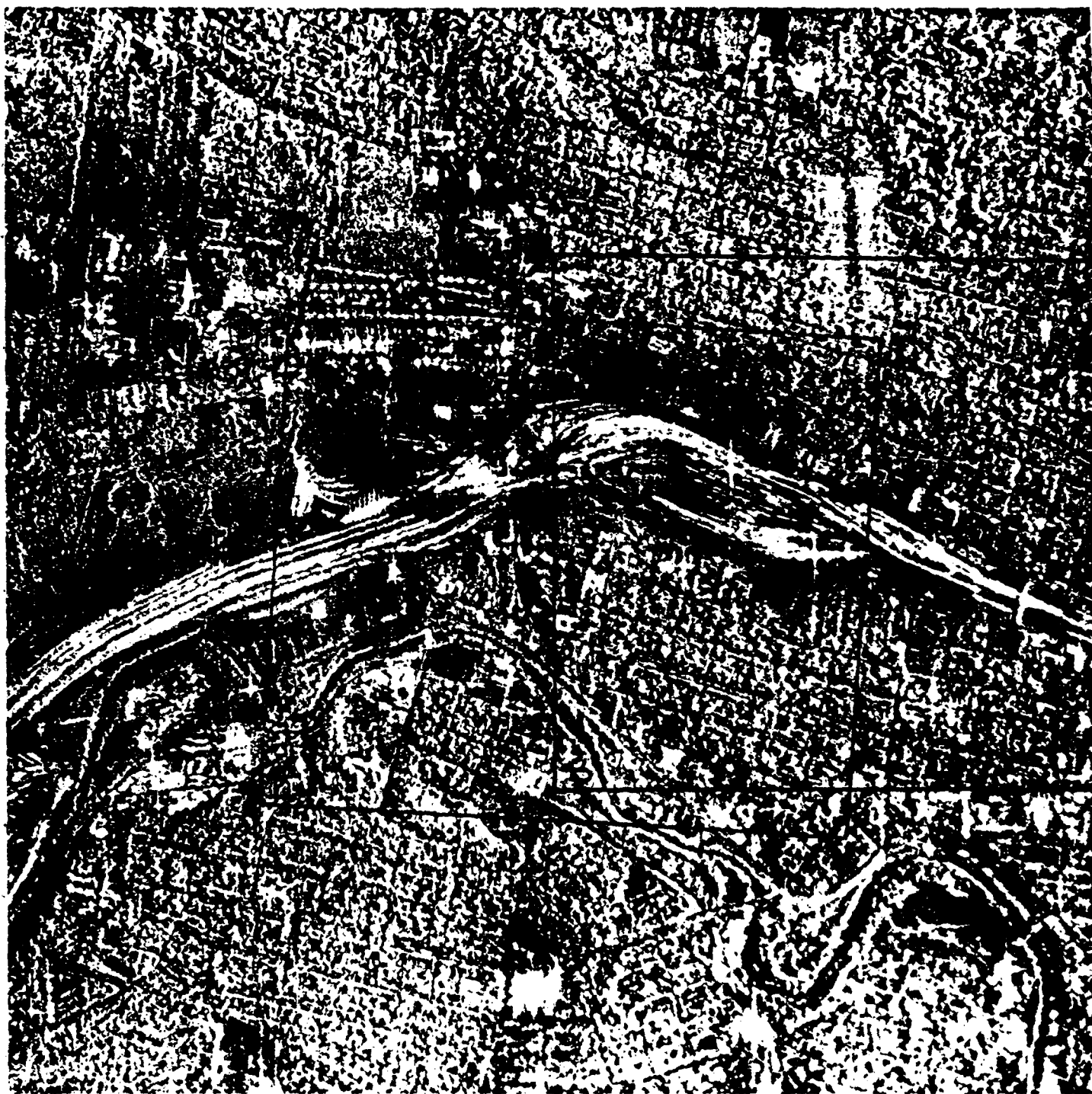


Figure 4

extensive features, (3) complex area features, and (4) individual man-made targets. Further each PI was asked to rank order each of the degraded images of each scene from best to worst, using both vehicle movement and activity level as criteria thereby providing a fifth response category. However, in all cases the ranking was identical for both applications. In addition, auxiliary data concerning the target area (e.g., maps, aerial photographs and large area coverage SAR imagery) were also provided. (See Appendix A for an interpreter package.)

In contrast to most image quality experiments, both absolute and relative (rankings) responses for the degraded images were obtained. This permitted the investigation of the correlation between the responses to various questions.

3.0 DEGRADATION PROCEDURES

The purpose of this section is to explain the processing steps that were performed on the digitized radar imagery to produce the desired set of experimental conditions. That is, the digitized radar images were processed to exhibit controlled levels of image quality. Thus for one digitized radar image several degraded images were generated and then evaluated by human interpreters. In this section the five processing algorithms, spatial frequency filtering, geometric distortion, noise addition, quantizing, and spatial domain averaging, which were used to degrade the radar images are presented.

A. Spatial Frequency Filtering

In the first processing step the digitized radar images were ideal low pass filtered in the spatial frequency domain. The purpose of this step was to limit the frequency content of the observed images.

B. Geometrical Distortion

Most image quality experiments do not consider geometric fidelity. A circularly symmetric geometric distortion was applied to the radar images. The distortion is defined by a simple sinusoidal compression, i.e.,

$$\Delta D = \sin \theta \Delta I \quad (3.1)$$

where: ΔI = a change in distance in the ideal (undistorted) image

ΔD = a change in distance in the distorted image,

and

$$\tan\theta = \tan\theta_1 + \frac{x}{w} \frac{\cos\theta_1 - \cos\theta_2}{\cos\theta_1 \cos\theta_2} \quad (3.2)$$

The two angles θ_1 and θ_2 are constants and vary the degree of geometric distortion. The distance to the center of the image from the edge is w and the distance from the center of the image to the center of ΔI is x . This type of distortion is similar to both the pin-cushion distortion and in radar, near range compression. By controlling the geometric fidelity of the degraded images it was hoped that a relationship between this image characteristic and image utility could be defined.

C. Additive Noise

Most image quality studies include the effect of additive white Gaussian noise. The third degradation applied to the original scenes was the addition of such noise. It is well known that for radar, fading noise is more significant than receiver noise. Receiver noise is usually modeled as being white additive while fading noise is neither white nor additive.

D. Number of Quantized Levels

An important parameter in system design, especially when digital processing is used or when the image data are transmitted over a communication channel, (as was the case with satellite data) is the number of quantized levels. This should be minimized while still maintaining a specified level of system performance. Therefore the number of quantized levels was varied to establish a relationship between radar image utility and the number of signal levels.

E. Spatial Domain Filtering

The final algorithm applied to the radar images was a simple uniform weighted square spatial domain filter. The purpose of this algorithm was to simulate various levels of system resolution.

4.0 RADAR IMAGE QUALITY METRICS

A set of five metrics was selected. Multiple metrics were used because it was hypothesized that no one metric could adequately characterize the utility of radar images. These metrics were intended to be independent measures of basic properties of radar imagery. There was one exception to this guideline, a root mean square error criterion was applied only because of its extensive use in other image analysis research.

Most of the quality metrics proposed in previous work can be related to one of these four metrics. For example, many researchers have proposed some measure of sharpness, resolution, edge quality, busyness, etc. All of these are directly related to the bandwidth of the image. Therefore, the first of our metrics was a root mean square bandwidth. Image quality has also been related to the dynamic range and signal-to-noise ratios; these image characteristics were also measured in this experiment. The last metric was specifically designed to estimate geometric fidelity. As mentioned previously geometric fidelity is an uncommon parameter to be treated in image quality studies, and as such a new metric had to be developed; this will be presented later. Each of the metrics was adapted for use on a digital computer.

In the following paragraphs the implementation of each of the metrics will be discussed.

A. Mean Square Bandwidth (MSB)

The definition of the MSB for a continuous one-dimensional signal, $f(t)$, is

$$\langle \omega^2 \rangle = \frac{\int_{-\infty}^{\infty} \omega^2 |F(\omega)|^2 d\omega}{\int_{-\infty}^{\infty} |F(\omega)|^2 d\omega} \quad (4.1)$$

where:

$$F(\omega) \Leftrightarrow [f(t)]$$

The fast Fourier transform (FFT) is required to obtain $F(\omega)$ for discrete signals. Defining $\hat{F}(i\Delta\omega)$ as the FFT of $f(t)$ then the MSB can be measured by evaluating

$$\langle \omega^2 \rangle = (\Delta\omega)^2 \frac{\sum_{i=1}^{N/2} (i-1)^2 |F(i\Delta\omega)|^2}{\sum_{i=1}^{N/2} |F(i\Delta\omega)|^2} \quad (4.2)$$

where:

$$\Delta\omega = \frac{1}{N\Delta t}$$

Δt = sample spacing

N = number of sample points

The actual metric used was $\sqrt{\langle \omega^2 \rangle}$, the root mean square bandwidth (RMSB). To use the RMSB as a valid indication of changes in image bandwidth for SAR imagery it must be assumed that the original imagery was fully focused. The RMSB is either directly or indirectly relatable to many previously proposed image quality metrics, e.g., sharpness. The above definition is only valid for one-dimensional signals. To extend its use to two-dimensions a further assumption was made, i.e., the MSB of a

two-dimensional image can be approximated by the product of the RMSB in orthogonal directions, e.g., along columns and rows of a digital image. This is a common assumption which is particularly valid for imaging radars (Harger, 1970).

B. Root Mean Square Error (RMSE)

The RMSE is commonly used as a quality criterion in image analysis research, especially in the bandwidth compression area. To allow for the comparison of the work with previous research the MSE as defined below was measured on each degraded image.

$$\epsilon^2 = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N (I(i,j) - D(i,j))^2 \quad (4.3)$$

where:

ϵ^2 = Mean Square Error

$N \times N$ = dimension of the image

$I(i,j)$ = pixel value at location i,j in the original SAR image

$D(i,j)$ = pixel value at location i,j in the degraded SAR image.

C. Image Dynamic Range

Dynamic range is a measure of the relationships between the "minimum" and "maximum" shades of grey (grey levels) in an images. There are different ways of defining both dynamic range and the minimum and maximum grey levels. Since we are working with digital imagery it is almost a certainty that at least one picture element (i.e., an outlying data point) will have the absolute minimum and another will have the absolute maximum of the allowable grey levels. A relative frequency approach is more suitable for our problem. That is, the minimum and maximum values would be deter-

mined a priori by establishing a fixed probability, for example, 5%, and measuring the highest and lowest grey levels which enclose 90% of the grey values. These grey levels would then become effective minimum and maximum values for the image.

Next we need to define how the effective values described above can be used to measure dynamic range or the contrast of the image. Five possible contrast measures are (Pratt, 1978)

$$\text{Contrast Ratio} = \frac{g_{\max}}{g_{\min}} \quad (4.4)$$

$$\text{Contrast Modulation} = \frac{g_{\max} - g_{\min}}{g_{\max} + g_{\min}} \quad (4.5)$$

$$\text{Differential Contrast} = \frac{g_{\max} - g_{\min}}{g_{\min}} \quad (4.6)$$

$$\text{Root Contrast Modulation} = \frac{\sqrt{g_{\max}} - \sqrt{g_{\min}}}{\sqrt{g_{\max}} + \sqrt{g_{\min}}} \quad (4.7)$$

and

$$\text{Relative Contrast} = \frac{g_{\max} - g_{\min}}{g_T} \quad (4.8)$$

where:

g_{\max} = effective maximum grey shade

g_{\min} = effective minimum grey shade

g_T = total possible range

Note that each of these contrast measures are dependent upon the same image characteristics, i.e., g_{\min} and g_{\max} ; thus, these are two measurements which are required. In our study we attempted to correlate relative

contrast with image utility.

D. Signal-to-Noise Ratio

Signal-to-noise ratio is an extremely important image characteristic, though its definition is entirely subject to the stochastic nature of the particular image formation process. We have already briefly discussed some aspects of photographic and radar image noise. Biberman (1974) analyzed the signal-to noise ratio for electro-optical imaging systems including the display. Any solely deterministic image quality analysis (i.e., not including random phenomena) can not provide an adequate assessment of the image because the probability of correctly identifying a target is strongly dependent upon the signal-to-noise ratio. Signal-to-noise ratios can be determined either by a rigorous theoretical analysis or by measurement; depending upon the intended application of the imagery either technique could suffice. For this study the signal-to-noise ratio is defined as

$$S/N = \bar{x}^2/S^2 \quad (4.9)$$

where S/N is the signal-to-noise ratio and \bar{x}^2/S^2 is the ratio of the square of the mean to the variance in "homogeneous" areas. A homogeneous area is a region in an image with stationary statistics, e.g., a wheat field.

E. A Geometric Fidelity Measure

The geometric fidelity of the degraded radar images was disturbed in this experiment as described previously. It was therefore necessary to develop an algorithm which would measure the geometric distortion introduced in these images. A heuristic approach was followed. This

approach was based on the realization that a geometric distortion is most visible on field boundaries and edges. Therefore, the RMS error between an edge image generated from the original data and an edge image derived from the degraded scene was used as a measure of geometric fidelity. The edge images were generated in both cases by a Robert's gradient. The Robert's gradient is defined as

$$G_r(j,k) = [F(j,k) - F(j+1,k+1)]^2 + [F(j,k+1) - F(j+1,k)]^2 \quad (4.10)$$

The edge images generated by a Robert's gradient are multi-level images and approximate a two-dimensional differentiation. If an edge in the degraded image is offset (distorted) with respect to the original scene, the RMS error will increase as the offset increases. The magnitude of the increase is dependent upon the edge contrast. This is the principle property any geometric fidelity measure must exhibit.

5.0 CORRELATION ANALYSIS

An important initial question that should be addressed before any models are estimated to predict image quality from interpreter responses is how much variation or how consistent were the interpreters in judging the degraded images. As a corollary to this initial question we are also interested in identifying individual interpreters whose responses vary significantly with respect to the other interpreters. Also as preparatory to estimating the regression models an analysis to assess how strongly the interpreter's responses are related to the image quality metrics is necessary. This information was obtained through a correlation analysis of the responses and measured data.

Define a matrix with 40 rows, one row corresponding to each degraded image, and M columns as y . A column in y will represent the interpreter responses, average interpreter response, degradation parameters, and quality metrics for each of the 40 degraded images. For example, if column i represented the fifth interpreter's responses and column j contains the average interpreter's responses then the correlation between the average interpreter and interpreter five, is defined as

$$r_{ij} = \frac{\sum_{k=1}^N n_{ki} n_{kj}}{(N-1) S_i S_j} \quad (5.1)$$

where:

$$S_i^2 = \frac{1}{N-1} \sum_{k=1}^N n_{ki}^2 \quad \text{and} \quad S_j^2 = \frac{1}{N-1} \sum_{k=1}^N n_{kj}^2 \quad (5.2)$$

where:

$$r_{ij} = y_{ij} - \frac{1}{N} \sum_{i=1}^N y_{ij} \quad (5.3)$$

The above simple correlation was used to analyze the relationships between the important elements of this study.

A. Analysis of Group and Individual Interpreters' Responses

An analysis was undertaken to determine the variation among the interpreter responses individually and in pairs. This was possible only because each interpreter provided five responses for each degraded image. Four of these responses related to the interpreter's judgment of his ability to identify, classify or detect targets in four target categories, (1) linear, (2) natural area features, (3) complex area features, and (4) individual man-made targets. These judgments are referred to as absolute responses. A fifth response category was obtained by having each interpreter rank the images from best to worst; these form a set of relative responses.

The variation among the interpreters' responses could thus be obtained by calculating r_{ij} between the responses of one interpreter for one category and another interpreter for the same category. Or the internal consistency of each interpreter could be evaluated by examining the r_{ij} for an individual interpreter across two categories. In this case if r_{ij} is near unity (e.g., greater than .8) then the interpreter's responses for the i^{th} and j^{th} categories were highly correlated and thus the interpreter was consistent in evaluating these categories. Note that low internal consistency indicates either that the interpreter did not judge the degraded images uniformly or that the criteria (i.e.,

the image information used) for judging one category is unrelated to the criteria used for judging the other category.

Table 1 contains the results of the internal consistency analysis. Obviously from this data the internal consistency of all eight interpreters and the average interpreter was very low. Thus either the interpreter did not judge the degraded images uniformly or the image information required for each category is different.

In Table 2 the correlation between each interpreter and the "average interpreter" is presented. If y_{ij} is the response for the j^{th} interpreter to the i^{th} image then the average interpreter response is defined as

$$\bar{y}_j = \frac{1}{M} \sum_{j=1}^M y_{ij} \quad (5.4)$$

where:

M = number of interpreters.

Consider the correlations for categories 1 and 5. The interpreters as a group were fairly consistent in responding to these categories as is evident by the magnitude of the correlation coefficient. Also, these correlation coefficients indicate that for category 1 interpreters 1, 4 and 6 were the most consistent, i.e., they were highly correlated with the average. The same interpreters are also strongly interrelated while the other five interpreters' responses were relatively uncorrelated. This is found by examining the correlation coefficient between individual interpreters. Similarly, interpreters 1, 2, 3, 7 and 8 are highly correlated for category 5. Table 2 also indicates that the interpreters as a group were very inconsistent in responding to categories 2, 3, and 4. This conclusion is also reached by examining the correlation between

TABLE 1
Correlation Coefficient for Each Interpreter

Category	Interpreter #								Ave.
	1	2	3	4	5	6	7	8	
1-2	.23	.44	.40	.23	.22	.2	.32	.16	.01
1-3	.34	.25	.50	.06	.18	.12	.15	.23	.11
1-4	.75	.42	.45	.38	.11	.08	.44	.32	.39
1-R	.18	.25	.19	.14	.03	.14	.47	.29	.26
2-3	.61	.09	.56	.03	.70	.31	.29	.52	.65
2-4	.40	.31	.11	.03	.54	.33	.06	.19	.30
2-R	.09	.45	.02	.00	.15	.22	.18	.22	.25
3-4	.38	.45	.05	.41	.61	.07	.51	.39	.49
3-R	.01	.07	.08	.40	.01	.45	.59	.39	.53
4-R	.01	.10	.24	.32	.04	.55	.68	.56	.61
Ave.	.30	.28	.26	.20	.26	.25	.37	.33	.36
Sd	.25	.16	.20	.17	.26	.16	.20	.14	.21

TABLE 2
Correlation of Each Interpreter With the
"Average Interpreter"

Category	PI							
	1	2	3	4	5	6	7	8
1	.95	.82	.83	.93	.76	.95	.84	.80
2	.67	.69	.50	.40	.47	.52	.52	.82
3	.29	.50	.18	.79	.61	.86	.72	.51
4	.50	.64	.51	.46	.02	.69	.70	.71
5	.94	.94	.82	.44	.78	.73	.81	.88

interpreters, i.e., for these categories the correlation coefficient between all pairs of interpreters was low.

From this correlation analysis we expect the regressions for categories 1 (linear features) and 5 (the relative ranking) to be superior to the estimated models for the other categories. Further, this analysis indicates that either the interpreters did not judge the degraded images uniformly or that the criteria for judging each category were unrelated.

B. Analysis of Interpreter Responses and Degradation Parameters and Image Quality Metrics

To support the regression analysis the correlation between each of the degradation parameters and image quality metrics and the interpreter responses was examined. This analysis provided an initial indication of the dependence of the image utility as a function of both the degradation parameters and quality metrics.

Table 3 contains the correlation coefficient between the average interpreter's responses and the parameters and metrics. Overall this correlation was low which indicates that the metrics and parameters alone do not provide in a linear sense a good indication of how the interpreters judged the degraded images. This reinforces our belief that a combination of image characteristics is required to predict the utility of radar images.

It is evident from the data that the parameters and metrics assume different levels of importance as the response category was changed. Consider the degradation parameters. All the parameters assumed approximately equal importance (as measured by the magnitude of the correlation coefficient) for categories 5 and 4, while only quantization showed any dependence for category 3. Similarly for the quality metrics, the S/N

TABLE 3

Correlation Between the Average Interpreter and
the Degradation Parameters and Quality Metrics

	Category				
	1	2	3	4	5
Degradation Parameters:					
Quantization	.01	.07	.40	.38	.63
Warp	.08	.13	.07	.41	.56
Bandwidth	.11	.34	.08	.46	.65
Spatial Filtering	.06	.27	.00	.42	.65
Noise	.15	.17	.16	.38	.57
Image Quality Metrics:					
RMSE	.21	.43	.17	.26	.46
RMSB	.30	.45	.17	.16	.27
Geometric Fidelity	.31	.40	.12	.24	.39
S/N	.74	.36	.23	.25	.09
Dynamic Range	.21	.22	.37	.40	.52

appears to be dominant for category 1 while it is completely uncorrelated for category 5. This is a significant observation in that it shows that image characteristics do not relate to image utility independent of the application of the data.

6.0 REGRESSION ANALYSIS

The culmination of our research is the prediction of radar image usefulness from measured image characteristics. In previous sections we have discussed the design of the experiment, the measured image characteristics and the correlation between these metrics and interpreter responses. The purpose of this section is to present a regression analysis which was performed to estimate the desired prediction equations. Several different approaches were pursued to establish adequate prediction equations. Obviously the first approach was to estimate a second order linear model from the observed data. A mixed model was used, i.e., a linear combination of image metrics, cross product terms (continuous variables) and blocking terms (discrete variables) were combined. The inclusion of blocking variables attempts to control undesirable fluctuations in responses occurring from different interpreters and scenes.

Another approach to finding an adequate prediction equation was to relate the interpreter response to the spatial-grey level volume. This metric has been proposed and applied to radar imagery in the past (Moore, 1979). A new measure was developed (based on the SGL volume) to incorporate the variation in dynamic range into the model. We call this metric the modified SGL volume. The models based on both the SGL and modified SGL volume are non-linear as will be explained later.

It is important to identify the criteria which will be applied to judge the usefulness of the estimated models. For linear models there are several well known criteria. For example, an F-statistic to test the significance of the model, the sum of squares for error (SSE), and the percentage of the total response variation explained by the model (R^2 , the coefficient of multiple determination) are all commonly used for

linear models; all of these criteria will be reported here for each estimated model; however, R^2 will be used primarily to evaluate the utility of the prediction equations. Unfortunately, it is difficult to judge the quality of estimated nonlinear models as will be required in evaluating the SGL based models. Here SSE (and estimated variance) will serve as the judging criteria for the nonlinear models.

The analysis of the linear model will be presented next; however, before the results are shown a short review will be included primarily to define terms. The SGL volume criteria will then be presented, followed by the estimated models.

6.1 The Linear Model and Results

6.1.1 Mathematical Basis

Because there are no specific physical laws governing the relationship between the utility of a microwave image and the image characteristics, a probabilistic model will be used to describe this relationship. Specifically, the general linear statistical model will be used. This model is written as

$$y = \beta_0 + \sum_{i=1}^k \beta_i z_i + \epsilon \quad (6.1)$$

where the z_i 's represent in this case known image parameters (or functions of parameters) and the β_i 's are unknown model parameters which define the desired relationship. This model is called "linear" because it is linear in the unknown model parameters. Both the β_i 's and z_i 's are deterministic values; the random component, ϵ , characterizes the stochastic nature of the observation y . The usual assumption for this model is that $E[\epsilon] = 0$ and

the $\text{Var} [\epsilon] = \sigma^2$ and that ϵ is normally distributed.

To reduce the number of observations required to estimate the β_i 's we will use the following specific form of the general linear model:

$$y = \gamma_0 + \sum_{i=1}^k \gamma_i x_i + \sum_{i=1}^k \gamma_{ii} x_i^2 + \sum_{j=1}^k \sum_{\substack{i=1 \\ i \neq j}}^k \gamma_{ij} x_i x_j \quad (6.2)$$

where:

$x_i = i^{\text{th}}$ image characteristic.

This model assumes that all three way interactions (e.g., $x_i \cdot x_j \cdot x_m$ terms) do not significantly contribute to the response y and thus are neglected. This model can also be viewed as a quadratic fit to the true higher order surface which defines the relationships between the image parameters and the response y .

6.1.2 Model Parameter Estimation

The model parameters are found by collecting observations and by performing a minimum mean square estimation. This procedure is well known but will be reviewed here.

Consider the following experiment. N observations of the utility of image data are obtained from M different radar images each processed to exhibit specific characteristics. Using the model presented in Section 6.1.1 the experiment can be mathematically written as

$$\underline{Y} = Z\beta + \epsilon \quad (6.3)$$

where:

$$\bar{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

$$Z = \begin{pmatrix} 1 & Z_{11} & Z_{\ell 1} \\ \vdots & \vdots & \vdots \\ 1 & Z_{1N} & Z_{\ell N} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{\ell} \end{pmatrix}$$

and

$$n = k^2 + k + 1$$

for the second order model described by Equation (6.2). Each ε_i is assumed to be zero mean Gaussian with identical variance σ^2 , further ε_i is independent of ε_j for all $i, j, i \neq j$. The matrix z defines the experimental conditions under which the observations were made and note that

$$\begin{aligned} \beta_i &= \gamma_m & \text{for } i = 0 \dots k, m = 0 \dots k \\ \beta_i &= \gamma_{m,m} & \text{for } i = k+1 \dots 2k+1, m = 0 \dots k \\ \beta_i &= \gamma_{m,n} & \text{for } i = 2k+2 \dots k^2+k+1, m = 1 \dots k, \\ & & n = 1 \dots k, m \neq n \end{aligned} \quad (6.4)$$

and

$$\begin{aligned} Z_i &= x_m & \text{for } i = 1 \dots k, m = 0 \dots k \\ Z_i &= x_m^2 & \text{for } i = k+1 \dots 2k+1, m = 0 \dots k \\ Z_i &= x_m x_n & \text{for } i = 2k+2 \dots k^2+k+1, m = 0 \dots k, \\ & & n = 0 \dots k, m \neq n \end{aligned} \quad (6.5)$$

The minimum mean square estimate for the model parameters, $\hat{\beta}$, is defined by the vector which minimizes, L defined as

$$L = (\underline{Y} - Z\hat{\beta})^T (\underline{Y} - Z\hat{\beta}) \quad (6.6)$$

Following (Meyers, 1977)

$$\begin{aligned} L &= (\underline{Y} - Z\hat{\beta})^T \underline{Y} - (\underline{Y} - Z\hat{\beta})^T Z\hat{\beta} \\ &= \underline{Y}^T \underline{Y} - (Z\hat{\beta})^T \underline{Y} - \underline{Y}^T Z\hat{\beta} + (Z\hat{\beta})^T Z\hat{\beta} \\ &= \underline{Y}^T \underline{Y} - \hat{\beta}^T Z^T \underline{Y} - \underline{Y}^T Z\hat{\beta} + \hat{\beta}^T Z^T Z\hat{\beta} \\ &= \underline{Y}^T \underline{Y} - 2\hat{\beta}^T Z^T \underline{Y} + \hat{\beta}^T Z^T Z\hat{\beta} \end{aligned} \quad (6.7)$$

Setting $\frac{\partial L}{\partial \hat{\beta}} = 0$ the best fit is found as

$$\frac{\partial L}{\partial \hat{\beta}} = -2Z^T \underline{Y} + 2Z^T Z\hat{\beta} = 0 \quad (6.8)$$

solving for

$$\hat{\beta} = (Z^T Z)^{-1} Z^T \underline{Y} \quad (6.9)$$

The expected value of $\hat{\beta}$ is simply found by

$$\begin{aligned} E[\hat{\beta}] &= E[(Z^T Z)^{-1} Z^T \underline{Y}] \\ &= E[(Z^T Z)^{-1} Z^T (Z\beta + \epsilon)] \\ &= E[(Z^T Z)^{-1} (Z^T Z\beta + Z^T \epsilon)] \\ &= E[(Z^T Z)^{-1} (Z^T Z)\beta + (Z^T Z)^{-1} Z^T \epsilon] \\ &= I\beta + E[(Z^T Z)^{-1} Z^T \epsilon] \\ &= \beta \end{aligned} \quad (6.10)$$

And the covariance matrix of $\hat{\beta}$ is found as (Meyers, 1977)

$$\begin{aligned}
 \text{Cov}[\hat{\beta}] &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] \\
 &= \text{Cov}[(Z^T Z)^{-1} Z^T \bar{Y}] \\
 &= (Z^T Z)^{-1} Z^T \text{Cov}(\bar{Y}) \\
 &= [(Z^T Z)^{-1} Z^T] \sigma^2 [I(Z^T Z)^{-1} Z^T]^T \\
 &= \sigma^2 (Z^T Z)^{-1}
 \end{aligned} \tag{6.11}$$

The minimum mean square estimate (MMSE) for the model parameters has been defined by the observation vector \bar{Y} and the design matrix Z . Further the MMSE was found to be unbiased and have a covariance $\sigma^2 (Z^T Z)^{-1}$. The covariance matrix can be used to establish confidence intervals for each of the model parameters and it allows the establishment of a prediction interval around any observation.

There are many ways in which an experiment can be designed (i.e., selection of the design matrix Z) to allow efficient estimation of the model parameters. Entire textbooks are devoted to presenting these techniques (Cox, 1958; Myers, 1977) for a wide variety of conditions. Therefore, a review of experimental design in general is not appropriate here.

6.1.3 Analysis of the Prediction Equation

The result of applying the technique described in the previous sections is a prediction equation which relates the image parameters to data utility for a specific application.

The data gathered in this experiment would also directly provide information concerning the magnitude of importance of an individual

parameter or groups of image parameters. Analysis of variance techniques are applied to obtain this information. For example, suppose we wish to test the hypothesis that image parameter x_1 does not significantly affect the utility of the sensor data. To investigate this question one would calculate the sum of squares for error (SSE) for the original model. The SSE is defined as

$$SSE = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (6.12)$$

where:

y_i = observed response for the experimental conditions defined by the i^{th} row in the design matrix Z .

\hat{y}_i = estimated (using eq. (6.9)) response for the experimental conditions defined by the i^{th} row in the design matrix Z .

Note that $SSE / (N - (\ell+1))$ is an estimate for σ^2 . Next a reduced model would be defined by deleting all x_1 terms in the original model. A SSE would then be calculated using the reduced model, SSE_1 . The test statistic for the given hypothesis is given by

$$F = \frac{\frac{SSE_1 - SSE}{\ell - g}}{\frac{SSE}{N - (\ell+1)}} \quad (6.13)$$

where:

N = number of observations

$\ell+1$ = number of parameters (β_i) in the original model

$g+1$ = number of parameters in the reduced model.

This test statistic has a F-distribution (probability density function) with $(\ell-g)$, $N-(\ell+1)$ degrees of freedom. Therefore, if $F > F_{(\ell-g), (N-(\ell+1)), \alpha}$ we reject the hypothesis that the sensor parameter x_1 does not significantly affect the utility of the sensor data with the probability of a TYPE I error equal to α . Remember that a TYPE I error consists of rejecting the hypothesis when it should be accepted. So if we reject, then the data has provided strong evidence that the image x_1 is important. Similar techniques can be applied to any term or group of terms in the original model. For example, we might wish to see if all quadratic terms, x_i^2 , do not significantly affect the response. The motivation would be to simplify the model and thus supply more degrees of freedom for estimating the remaining parameters. If only one parameter, z_i , is tested then the above F value is called a partial F-test value. An overall F-test for the regression is found by testing the hypothesis that $\beta_1 = \dots = \beta_\ell = 0$. If this F-test statistic exceeds some value specified from a selected risk level α then the regression is said to be statistically significant. That is, the variation of the data predicted by the model is greater than would be expected by a random occurrence at a probability of $1-\alpha$. Although a specific estimated regression equation is statistically significant it does not follow necessarily that this equation is useful for prediction. One rule of thumb has been developed (Wetz, 1964) which states that if the F-test statistic is four times greater than $F_{k, (N-(\ell+1)), \alpha}$ then the regression equation in question is satisfactory for prediction.

The coefficient of multiple determination, R^2 , (Draper and Smith, 1966) is defined as

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (6.14)$$

where:

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \quad (6.15)$$

This quantity represents the percentage of the variation in the response data explained by the estimated model. Note that if $N = 2$, then $R^2 = 1$ because there is a perfect fit to the data; therefore, when using R^2 as a goodness criterion care must be taken to insure that there are sufficient degrees of freedom.

Sometimes it is convenient to establish a prediction interval (Mendenhall, 1968) around a specific response. This interval would define a range in which some future response would lie given this set of parameters and thus provide an indication of the operational utility of the images derived from a specific system. To develop the prediction interval we will first define the set of parameters as the vector 0 , i.e.,

$$0^T = (1, x_{0_1}, \dots, x_{0_k}) \quad (6.16)$$

The expected value of the response given the vector 0 is

$$E[\hat{Y}] = 0^T \beta \quad (6.17)$$

and its variance can be shown to be

$$\text{Var}[\hat{Y}] = [0^T (Z^T Z)^{-1} 0] \sigma^2 \quad (6.18)$$

The error, E , between a future response, y_F , and the estimated response, \hat{y} , as defined by the estimated model parameters is

$$E = y_F - \hat{y} \quad (6.19)$$

so clearly

$$E[E] = 0 \quad (6.20)$$

$$\text{Var}[E] = \text{Var}[y_F] + \text{Var}[\hat{y}] - 2 \text{Cov}(y_F, \hat{y}) \quad (6.21)$$

but the future and the estimated response would be uncorrelated yielding

$$\text{Var}[E] = \text{Var}[y] + \text{Var}[y_F] \quad (6.22)$$

The variance of the future utility of the image data, y_F , is assumed to be σ^2 , thus

$$\text{Var}[E] = \sigma^2 [1 + 0^T (Z^T Z)^{-1} 0] \quad (6.23)$$

A prediction interval, $(1 - \alpha)\%$, can now be defined (remember y and y_F are both normal so E is also normal) as

$$\hat{y} \pm t_{N-(k+1), \alpha/2} S \sqrt{1 + 0^T (Z^T Z)^{-1} 0} \quad (6.24)$$

where:

$$S^2 = \text{estimate of } \sigma^2 = \text{SSE}/N-(k+1)$$

$$t_{N-(k+1), \alpha/2} = \text{the } t \text{ value for } N-(k+1) \text{ degrees of freedom at } \alpha/2.$$

The interpretation of this interval is simple; there is a $(1 - \alpha)\%$ probability that a future measurement of the sensor's data utility will lie inside this interval.

6.1.4 Blocking

Additional variation in the response could be introduced into this experiment from differences between interpreters and differences between the

four scenes. A technique to reduce these variations is thus required. A method known as blocking was employed to accomplish this task. Blocking will be explained by the following example.

Suppose we conducted this experiment with only three interpreters and two scenes; further, assume that only one variable (image property) is used. The following linear statistical model is proposed:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon \quad (6.25)$$

where:

y = interpreter's response

x_1 = image variable

$x_2 = 1$, if the response is from interpreter #1

$x_2 = 0$, if the response is not from interpreter #1

$x_3 = 1$, if the response is from interpreter #2

$x_3 = 0$, if the response is not from interpreter #2

$x_4 = 1$, if the response is from scene A

$x_4 = 0$, if the response is not from scene A

ϵ = random error

Note that $x_2 = x_3 = 0$ implies that the response is from interpreter #3, similarly $x_4 = 0$ implies that the response is from scene B. Next consider the interpretation of β_2 , β_3 and β_4 . Under the assumption that interpreter #1 and scene B was used, the model takes the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \quad (6.26)$$

while the assumption that interpreter #3 and scene B was used yields

$$y = \beta_0 + \beta_1 x_1 + \epsilon. \quad (6.27)$$

Clearly the parameter β_2 represents the amount that we expect the response to increase or decrease on the average as we move from interpreter #3 to interpreter #1. Similarly β_3 represents the expected difference between interpreter #2 and #1, while β_4 is the expected difference between scene A and B.

Next consider three specific observations.

$$y_{1B} = \beta_0 + \beta_1 x_1 + \epsilon_{1B} \quad (\text{from interpreter \#3, scene B}) \quad (6.28)$$

$$y_{2B} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon_{2B} \quad (\text{from interpreter \#1, scene B}) \quad (6.29)$$

$$y_{3B} = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \epsilon_{3B} \quad (\text{from interpreter \#2, scene B}) \quad (6.30)$$

Averaging over the three interpreters we obtain:

$$\bar{y}_B = \frac{1}{3} \sum_{i=1}^3 y_{iB} = \beta_0 + \beta_1 x_1 + \frac{\beta_2 + \beta_3}{3} + \bar{\epsilon}_B \quad (6.31)$$

where:

$$\bar{\epsilon}_B = \frac{1}{3} \sum_{i=1}^3 \epsilon_{iB} \quad (6.32)$$

Similarly for scene A

$$y_{1A} = \beta_0 + \beta_1 x_1 + \beta_4 x_4 + \epsilon_{1A} \quad (6.33)$$

$$y_{2A} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_4 x_4 + \epsilon_{2A} \quad (6.34)$$

$$y_{3A} = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \beta_4 x_4 + \epsilon_{3A} \quad (6.35)$$

and

$$\bar{y}_A = \beta_0 + \beta_1 x_1 + \beta_4 x_4 + \frac{\beta_2 + \beta_3}{3} + \bar{\epsilon}_A \quad (6.36)$$

To estimate the difference between the responses from scene A and B we form

$$\bar{y}_A - \bar{y}_B = \beta_4 + (\bar{\epsilon}_A - \bar{\epsilon}_B) \quad (6.37)$$

where:

$$\bar{\epsilon}_A - \bar{\epsilon}_B = \text{error of estimation.}$$

The effects of β_0 , β_1 and $\beta_2 + \beta_3/2$ cancel out thereby reducing the error in estimating β_4 . Similar observations can be made concerning β_2 and β_3 .

Blocking reduces the error associated with estimating the model parameters associated the desired experimental variables caused by fluctuations from different interpretations and scenes. It also allows for the investigation of the magnitude of these fluctuations.

6.1.5 Results of the Linear Regression Analysis

A full model with 21 parameters defined by equation (6.2) and 10 blocking parameters was proposed and the minimum mean square estimation for parameters found. Of the 10 blocking parameters, 7 account for the eight interpreters, and 3 account for the four different scenes. A reduced model (one in which no blocking variables were incorporated) was also used. The purpose of this section is to present the results of the regression analysis and to indicate appropriate conclusions which may be drawn from it.

Table 4 contains the coefficient of multiple determination, R^2 , SSE, the F-test value, and the estimated random variance for the full and reduced models for all five response categories. Also included in Table 4 are results from reduced models where the cross terms, $x_i x_j$, $i \neq j$, have been deleted and where the square terms x_i^2 and cross terms have been removed.

The first observation to be made from this data is that the predic-

Table 4
Evaluation of Predicted Regression Models

	Category				
	1	2	3	4	5
Full Model					
R^2	76.7	54.0	64.7	52.6	56.0
F	30.9	9.45	17.2	10.7	12.2
# Degrees of Freedom	30, 281	30, 241	30, 281	30, 289	30, 289
$F_{k,N-(k+1),.05}$	1.52	1.52	1.52	1.52	1.52
S^2	1.0	1.4	.8	1.1	.04
SSE	285.2	346.6	225.0	328.3	11.66
Reduced Model (no blocking)					
R^2	58.9	17.0	9.8	15.2	53.8
F	20.9	2.6	1.6	2.7	17.4
# Degrees of Freedom	20, 291	20, 251	20, 291	20, 299	20, 299
$F_{k,N-(k+1),\alpha}$	1.62	1.62	1.62	1.62	1.62
S^2	1.7	2.5	2.0	2.0	.04
SEE	503.7	626.2	575.5	586.9	12.23
Reduced Model (no blocking and no $x_i x_j$ terms)					
R^2	52.0				45.5
F	32.6				25.8
# Degrees of Freedom	10, 301				10, 309
$F_{k,N-(k+1),\alpha}$	1.85				1.85
S^2	1.9				.047
SSE	588.6				14.2
Reduced Model (no blocking and no x^2 , and $x_i x_j$ terms)					
R^2	40.9				36.5
F	42.4				36.1
# Degrees of Freedom	5, 306				5, 314
$F_{k,N-(k+1),\alpha}$	2.23				2.23
S^2	2.4				.053
SSE	724.6				16.8

tion equation is statistically significant for all response categories when the blocking variables are included in the model. Further review shows that response category 1 (linear features) provided the best prediction equation in terms of maximum R^2 .

However, when the blocking variables are removed only response categories 1 and 5 still produce statistically significant regression equations. Obviously most of the variation in the interpreter responses for categories 2, 3 and 4 predicted by the full model was accounted for in the blocking variables and not by the image metrics. This was expected from the results of the correlation analysis, i.e., the correlation between interpreters was low for response categories 2, 3, and 4. The interpreters were not consistent in judging these categories and therefore a reasonable prediction equation can not be estimated. These results also indicate that reasonable prediction equations can be obtained for response categories 1 and 5, so the rest of our discussion will only deal with these categories. Note that this result could also be predicted from the correlation analysis because the correlation between interpreters was reasonably high for categories 1 and 5 and, therefore, the interpreters were consistent in judging the images with respect to these response categories.

Conducting a hypothesis test to determine if the blocking variables are needed we find that for category 1

$$F = \frac{\frac{SEE_1 - SSE}{2-g}}{\frac{SSE}{N-(g+1)}} = \frac{503.7 - 285.2}{10} \div \frac{285.2}{281} = 21.53 \quad (6.30)$$

but $F_{(k-g), (N-(l+1)), .05} = F_{10, 281, .05} \approx 1.88$. Therefore we can not reject the hypothesis that the blocking variables are not required. This indicates that there was a statistically significant difference between the responses for various interpreters and scenes. For response category 5 the F-test value is 1.39 while $F_{10, 299, .05} \approx 1.88$, therefore we can accept the hypothesis that the blocking variables are not required for the relative ranking of the degraded images. Note that this standard hypothesis test is constructed to provide information when the hypothesis is rejected as with category 1. However, given the magnitude of the F-test value we can make some comments as to how strongly we believe that the hypothesis is truly correct. This is done by computing the probability that $f < F$ where f has a F-pdf with 10 and 299 degrees of freedom. If this probability is close to $1-\alpha$ then we would not feel strongly that the hypothesis is correct; however, if $P[f < F]$ is far from $1-\alpha$ then it would be reasonable to accept the hypothesis. For this case $p = P[f < F] = .81$. We feel that this is strong enough evidence to indicate that the blocking variables are not required for category 5. Using the same approach we found that both the cross terms $x_i x_j$ and the squared term x_i^2 are required in the model for both response categories.

Now that we have established the quality of the prediction equation it is left to present the coefficient estimates and discuss the relative importance of various model parameters. Four models, i.e., prediction equations, will be presented for response categories 1 and 5.

Table 5 contains the F-test value for ten reduced models. These results were generated to establish the relative importance of each of the five metrics. The first five of the models contained the blocking variables while the variables (linear, squared, and cross) associated with

Table 5

Evaluation of the Importance of Image Quality Parameters

	Category	
	1	5
Reduced Model Variable Terms Removed		
RMSE		
R^2	76.26	53.56
F	.99	2.67
RMSB		
R^2	74.8	52.3
F	3.93	3.92
Geo		
R^2	76.26	52.4
F	.98	3.96
S/N		
R^2	75.2	51.8
F	3.13	2.25
DYN		
R^2	76.3	38.6
F	.92	19.21
Blocking and Variable Terms Removed		
RMSE		
R^2	47.7	50.3
F	13.5	3.88
RMSB		
R^2	54.7	50.7
F	5.09	3.63
Geo		
R^2	48.4	49.6
F	12.63	6.83
S/N		
R^2	55.0	48.6
F	4.73	6.17
DYN		
R^2	51.1	33.8
F	9.39	22.17

each of the metrics was deleted. The second five models deleted both the blocking variables and the specific metric variables. The higher the F-test value for a particular model the more significant the deleted variables are for predicting the response.

As indicated in the correlation analysis the order of importance of the five metrics is different for different response categories. Considering the models with the blocking variables included we notice that the RMSB and S/N have about equal importance, while with the other three we accept the hypothesis that they are not required taken individually in the model with a $p = .56$ for category 1. For category 5 all the metrics are required in the model but the dynamic range is by far the most significant with the other four being about of equal importance. Interestingly, when the blocking variables are removed the order of importance of the metrics is changed for category 1. Now the RMSB and S/N are the least important metrics for category 1. However, because the blocking variables are required for category 1 the initial ordering of the importance of the metrics will be accepted.

Tables 6-13 contain the estimated regression coefficient estimates for several different models for category 1 and 5. These tables include upper and lower confidence limits ($\alpha=.1$) for each coefficient estimate, its standard error, adjusted sum of squares and its partial F-test value. These tables represent image quality predictions based on measured image properties.

From this regression analysis several conclusions can be stated:

- a) Statistically significant regression equations can not be estimated for response categories 2, 3, and 4.
- b) Statistically significant regression equations can be estimated

TABLE 6

Regression Coefficients for Response Category 1 for the Full Model

β_i	Lower Confidence Limit	Estimated Regression Coefficient	Upper Confidence Limit	Standard Error	Adjusted Sum of Squares	Partial F-test Value
Linear terms						
x_1	1	-17.01012	18.47463	10.75121	0.00471	0.00464
	2	-7.33279	9.89595	5.21998	0.06119	0.06028
	3	-27.40499	14.48761	12.69266	0.26286	0.25893
	4	-7.67329	10.71162	5.57028	0.07551	0.07438
	5	-5.58449	29.48669	10.62590	1.28417	1.26498
Squared terms						
x_i^2	6	-14.48670	54.17187	20.80225	0.92367	0.90986
	7	-5.43830	11.43042	5.11090	0.34885	0.34364
	8	-39.95377	47.98805	26.64471	0.02308	0.02273
	9	-4.89679	5.54673	3.16419	0.01071	0.01055
	10	-19.86338	21.00693	12.38293	0.00216	0.00213
Cross terms						
$x_i x_j$	11	-116.80167	28.05670	43.88934	1.03764	1.02214
	12	-87.16297	55.56056	43.24253	0.13555	0.13352
	13	-32.43962	4.43666	11.17282	1.59427	1.57044
	14	-57.16200	30.55782	26.57744	0.25430	0.25050
	15	-20.23500	106.86131	38.50778	1.28435	1.26515
	16	-72.51422	8.50954	24.54867	1.72523	1.69945
	17	-47.28237	8.38061	16.86483	1.35037	1.33019
	18	-15.54164	7.35019	13.87158	0.28503	0.28077
	19	-29.68441	1.67420	19.00213	0.00788	0.00776
	20	-19.22218	-3.56538	9.48712	0.14342	0.14127
Blocking Terms						
	21	-0.29192	0.08462	0.22817	0.13962	0.13753
	22	-1.61269	-1.23615	0.22817	29.79748	29.35217
	23	-1.59013	-1.21359	0.22817	28.71959	28.29039
	24	-0.86731	-0.49077	0.22817	4.69666	4.62647
	25	-2.81884	-2.44231	0.22817	116.31489	114.57660
	26	-0.94243	-0.56590	0.22817	6.24467	6.15135
	27	-1.13782	-0.76128	0.22817	11.30123	11.13233
	28	3.75469	6.14191	1.44656	18.30085	18.02735
	29	-0.32245	1.19997	0.92253	1.71758	1.69192
	30	2.59943	4.34048	1.05501	17.18300	16.92621
	0	-1.36198	1.95391	2.00931	0.	0.

 x_1 = RMSE x_2 = RMSB x_3 = Geometric Fidelity x_4 = S/N x_5 = Dynamic Range x_{21} = PI-1, i.e., $x_{21}=1$ if response is from PI-1

:

 x_{27} = PI-7 x_{28} = Scene 1 x_{29} = Scene 2 x_{30} = Scene 3

TABLE 7

Regression Coefficients for Response Category 1 for the Reduced Model (No Blocking)

	β_i	Lower Confidence Limit	Estimated Regression Coefficient	Upper Confidence Limit	Standard Error	Adjusted Sum of Squares	Partial F-test Value
Linear terms x_1	1	-10.60412	4.11201	18.82813	8.91083	0.36802	0.21260
	2	-14.73150	-6.89804	0.93541	4.74710	3.65510	2.11152
	3	-10.76443	2.96381	16.69204	8.31936	0.21970	0.12692
	4	-13.44627	-6.60710	0.23206	4.14456	4.39916	2.54135
	5	-7.98058	8.05794	24.09646	9.71941	1.18980	0.68734
Squared terms x_i^2	6	73.29022	103.65431	134.01839	18.40076	54.92986	31.73242
	7	0.87286	11.28396	21.69506	6.30917	5.53712	3.19874
	8	-3.78126	33.47656	70.73438	22.57839	3.80541	2.19834
	9	-2.57430	3.31579	9.20587	3.56941	1.49377	0.86294
	10	-55.28663	-38.68179	-22.07695	10.06260	25.57985	14.77722
Cross terms $x_i x_j$	11	-280.21721	-208.14718	-136.07715	43.67473	39.31751	22.71332
	12	-202.94308	-136.52495	-70.10681	40.24966	19.91613	11.50535
	13	-51.73406	-32.15542	-12.57678	11.86474	12.71445	7.34501
	14	-21.31386	19.92495	61.16377	24.99089	1.10036	0.63567
	15	116.61236	181.31228	246.01219	39.20841	37.01701	21.38435
	16	7.94868	38.25996	68.57123	18.36876	7.50991	4.33840
	17	-10.57893	12.57210	35.72313	14.02962	1.39005	0.80302
	18	3.44209	27.98858	52.53507	14.87527	6.12826	3.54023
	19	-46.85694	-13.00060	20.85574	20.51708	0.69503	0.40151
	20	7.06207	20.58972	34.11737	8.19781	10.91969	6.30820
	0	-3.33727	-0.30112	2.73503	1.83992	0.	0.

 x_1 = RMSE x_2 = RMSB x_3 = Geometric Fidelity x_4 = S/N x_5 = Dynamic Range x_{21} = PI-1, i.e., $x_{21}=1$ if response is from PI-1

:

 x_{27} = PI-7 x_{28} = Scene 1 x_{29} = Scene 2 x_{30} = Scene 3

TABLE 8
Regression Coefficients for Response Category 1 for the Reduced Model
(No Blocking and No $x_i x_j$ terms)

	β_i	Lower Confidence Limit	Estimated Regression Coefficient	Upper Confidence Limit	Standard Error	Adjusted Sum of Squares	Partial F-test Value
Linear terms x_i	1	-26.86310	-19.31937	-11.77564	4.57200	34.91623	17.85552
	2	-5.69371	-2.10589	1.48192	2.17446	1.83412	0.93793
	3	4.35229	12.02003	19.68776	4.64716	13.08249	6.69014
	4	4.17893	6.61806	9.05719	1.47828	39.19262	20.04238
	5	9.88061	15.10121	20.32181	3.16403	44.54471	22.77934
Squared terms x_i^2	6	17.03603	23.98216	30.92828	4.20982	63.46070	32.45263
	7	-1.73761	1.63879	5.01519	2.04632	1.25415	0.64135
	8	-23.74625	-17.09444	-10.44264	4.03144	35.15967	17.98001
	9	-6.06277	-3.80679	-1.55080	1.36728	15.15852	7.75179
	10	-34.00553	-25.81383	-17.62212	4.96472	52.86517	27.03428
	0	-0.11491	1.10453	2.32397	0.73906	0.	0.

x_1 = RMSE

x_2 = RMSB

x_3 = Geometric Fidelity

x_4 = S/N

x_5 = Dynamic Range

x_{21} = PI-1, i.e., $x_{21}=1$ if response is from PI-1

⋮

x_{27} = PI-7

x_{28} = Scene 1

x_{29} = Scene 2

x_{30} = Scene 3

TABLE 9
Regression Coefficients for Response Category 1₂ for the Reduced Model
(No Blocking and No $x_i x_j$ or x_i^2 terms)

	β_i	Lower Confidence Limit	Estimated Regression Coefficient	Upper Confidence Limit	Standard Error	Adjusted Sum of Squares	Partial F-test Value
Linear terms x_i	1	2.33357	4.53349	6.73342	1.33343	27.37395	11.55912
	2	-1.28237	-0.41925	0.44387	0.52316	1.52089	0.64222
	3	-7.66419	-5.58307	-3.50194	1.26142	46.39132	19.58953
	4	3.33313	3.96515	4.59716	0.38308	253.71889	107.13712
	5	0.20011	1.39658	2.59306	0.72522	8.78237	3.70851
	0	0.86941	1.42709	1.98476	0.33802	0.	0.

x_1 = RMSE
 x_2 = RMSB
 x_3 = Geometric Fidelity
 x_4 = S/N
 x_5 = Dynamic Range
 x_{21} = PI-1, i.e., $x_{21}=1$ if response is from PI-1
 \vdots
 x_{27} = PI-7
 x_{28} = Scene 1
 x_{29} = Scene 2
 x_{30} = Scene 3

TABLE 10

Regression Coefficients for Response Category 5 for the Full Model

β_i	Lower Confidence Limit	Estimated Regression Coefficient	Upper Confidence Limit	Standard Error	Adjusted Sum of Squares	Partial F-test Value
Linear terms						
x_1	1	0.76659	4.12928	2.03787	0.16573	4.10577
	2	-2.11382	-0.39908	1.03917	0.00595	0.14749
	3	-8.57619	-4.97604	2.18177	0.20997	5.20174
	4	-1.92364	-0.09097	1.11064	0.00027	0.00671
	5	2.72216	5.85721	1.89992	0.38364	9.50414
Squared terms						
	6	-10.66498	-3.84136	4.13528	0.03483	0.86290
	7	-0.91857	0.72126	0.99377	0.02126	0.52675
	8	-8.34392	-0.83348	4.55151	0.00135	0.03353
	9	-1.74720	-0.75447	0.60162	0.06348	1.57269
	10	-6.09660	-2.35796	2.26571	0.04372	1.08309
Cross terms						
	11	-18.25007	-3.81583	8.74750	0.00768	0.19029
	12	-7.87051	5.71423	8.23268	0.01945	0.48176
	13	-6.76284	-3.15659	2.18548	0.08421	2.08614
	14	-19.79533	-12.65374	4.32798	0.34505	8.54807
	15	-8.60601	4.03993	7.66374	0.01122	0.27789
	16	-13.04644	-4.96841	4.89548	0.04158	1.03002
	17	-8.22106	-2.73964	3.32187	0.02746	0.68018
	18	-0.42680	3.88487	2.61297	0.08923	2.21046
	19	5.19638	9.34536	2.51438	0.55763	13.81436
	20	-4.06203	-0.95051	1.88566	0.01026	0.25409
Blocking Terms						
	21	-0.07622	-0.00209	0.04493	0.00009	0.00216
	22	-0.07872	-0.00459	0.04493	0.00042	0.01044
	23	-0.07205	0.00208	0.07622	0.00009	0.00215
	24	-0.07413	-0.00000	0.07413	0.00000	0.00000
	25	-0.07413	-0.00000	0.07413	0.00000	0.00000
	26	-0.07413	-0.00000	0.07413	0.00000	0.00000
	27	-0.07613	-0.00200	0.07213	0.00008	0.00198
	28	0.38091	0.84119	1.30146	0.36710	9.09421
	29	0.33572	0.60305	0.87037	0.55931	13.85610
	30	0.31736	0.63875	0.96014	0.43414	10.75510
	0	-0.82954	-0.20455	0.42044	0.37876	0.

 x_1 = RMSE x_2 = RMSB x_3 = Geometric Fidelity x_4 = S/N x_5 = Dynamic Range x_{21} = PI-1, i.e., $x_{21}=1$ if

response is from PI-1

:

 x_{27} = PI-7 x_{28} = Scene 1 x_{29} = Scene 2 x_{30} = Scene 3

TABLE 11
Regression Coefficients for Response Category 5 Reduced Model (No Blocking)

β_i	Lower Confidence Limit	Estimated Regression Coefficient	Upper Confidence Limit	Standard Error	Adjusted Sum of Squares	Partial F-test Value
Linear terms						
1	-3.42016	-1.16063	1.09889	1.36944	0.02939	0.71829
2	-1.55801	-0.37888	0.80025	0.71464	0.01150	0.28108
3	-0.41467	1.64799	3.71065	1.25013	0.07111	1.73779
4	-2.04706	-1.01073	0.02560	0.62810	0.10597	2.58952
5	3.36761	5.68876	8.00991	1.40680	0.66915	16.35205
Squared terms						
6	-6.01712	-1.34948	3.31816	2.82895	0.00931	0.22755
7	-0.77492	0.75875	2.29242	0.92952	0.02727	0.66631
8	-13.83206	-8.32026	-2.80847	3.34057	0.25385	6.20344
9	-1.12554	-0.27170	0.58214	0.51749	0.01128	0.27566
10	-9.17112	-6.84725	-4.52337	1.40845	0.96716	23.63480
Cross terms						
11	-16.91121	-5.84652	5.21816	6.70605	0.03110	0.76008
12	-0.63428	9.51973	19.67374	6.15411	0.09792	2.39287
13	-3.28568	-0.27793	2.72981	1.82292	0.00095	0.02325
14	-16.66335	-11.31343	-5.96351	3.24246	0.49818	12.17415
15	-5.21931	4.72386	14.66702	6.02632	0.02514	0.61445
16	-5.57293	-0.98884	3.59525	2.77831	0.00518	0.12667
17	-3.78832	-0.23351	3.32130	2.15449	0.00048	0.01175
18	-2.94127	0.80658	4.55443	2.27149	0.00516	0.12609
19	5.71524	9.48723	13.25923	2.28612	0.70474	17.22190
20	0.36464	2.40731	4.44997	1.23801	0.15473	3.78105
0	-0.36411	0.07813	0.52036	0.26803	0.	0.

x_1 = RMSE

x_2 = RMSB

x_3 = Geometric Fidelity

x_4 = S/N

x_5 = Dynamic Range

x_{21} = PI-1, i.e., $x_{21}=1$ if

response is from PI-1

:

x_{27} = PI-7

x_{28} = Scene 1

x_{29} = Scene 2

x_{30} = Scene 3

TABLE 12
Regression Coefficients for Response Category 5 Reduced Model
(No Blocking or $x_i x_j$ terms)

	β_i	Lower Confidence Limit	Estimated Regression Coefficient	Upper Confidence Limit	Standard Error	Adjusted Sum of Squares	Partial F-test Value
Linear terms x_i	1	-1.32808	-0.21548	0.89712	0.67441	0.00477	0.10208
	2	-1.00396	-0.45234	0.09928	0.33437	0.08546	1.83011
	3	-1.41974	-0.32430	0.77114	0.66401	0.01114	0.23853
	4	-0.27638	0.08341	0.44320	0.21809	0.00683	0.14628
	5	2.73778	3.46737	4.19695	0.44224	2.87059	61.47194
Squared terms x_i^2	6	-1.30139	-0.29636	0.70866	0.60921	0.01105	0.23666
	7	-0.15507	0.36478	0.88462	0.31511	0.06258	1.34008
	8	-0.64720	0.26369	1.17458	0.55214	0.01065	0.22808
	9	-0.53077	-0.19404	0.14269	0.20411	0.04220	0.90377
	10	-5.10945	-3.92660	-2.74375	0.71700	1.40053	29.99144
	0	0.20974	0.39093	0.57212	0.10983	0.	0.

x_1 = RMSE

x_2 = RMSB

x_3 = Geometric Fidelity

x_4 = S/N

x_5 = Dynamic Range

x_{21} = PI-1, i.e., $x_{21}=1$ if response is from PI-1

:

x_{27} = PI-7

x_{28} = Scene 1

x_{29} = Scene 2

x_{30} = Scene 3

TABLE 13
Regression Coefficients for Response Category 5 Reduced Model
(No Blocking, $x_i x_j$ or x_i^2 terms)

β_i	Lower Confidence Limit	Estimated Regression Coefficient	Upper Confidence Limit	Standard Error	Adjusted Sum of Squares	Partial F-test Value
Linear terms						
x_1	-1.17260	-0.85077	-0.52893	0.19508	1.01904	19.01985
2	-0.27588	-0.14611	-0.01634	0.07866	0.18487	3.45044
3	0.04973	0.34977	0.64981	0.18187	0.19818	3.69887
4	-0.08228	0.01243	0.10713	0.05740	0.00251	0.04688
5	0.85605	1.03178	1.20751	0.10652	5.02688	93.82428
0	0.52412	0.60738	0.69064	0.05047	0.	0.

x_1 = RMSE

x_2 = RMSB

x_3 = Geometric Fidelity

x_4 = S/N

x_5 = Dynamic Range

x_{21} = PI-1, i.e., $x_{21}=1$ if
response is from PI-1

:

x_{27} = PI-7

x_{28} = Scene 1

x_{29} = Scene 2

x_{30} = Scene 3

for response categories 1 and 5 even when the models are reduced.

- c) The blocking variables are required for response category 1 but not for response category 5.
- d) The square and cross product terms are required for both response category 1 and 5.
- e) RMSB and S/N are the most important variables for response category 1.
- f) Dynamic range is the most important variable for response category 5.

6.2 Non-Linear Models and Results

Recently a study was conducted (Moore, 1979) which related the "interpretability" of radar images to the product of the spatial resolution and grey-level resolution. The purpose of this section is to determine if this relationship can be derived from the experimental data used here.

The product of the spatial resolution and the grey-level resolution is defined as the spatial-grey-level (SGL) volume. The nonlinear model proposed in Moore (1979) to relate interpretability to the SGL was

$$I = I_0 \exp \left[-\frac{V}{V_c} \right] \quad (6.39)$$

where:

I_0 and V_c are the model parameters

$V = r_s r_{gL}$

r_s = spatial resolution (2-dimensional)

r_{gL} = grey-level resolution

The grey-level resolution is a measure of the width of the noise probability density function (pdf) for a specific system, i.e., the noise variance is dependent upon the amount of averaging performed by the system. The measure used in Moore (1979) to measure the width of the pdf was the ratio of 90% to the 10% points. For a Gaussian approximation of the gamma pdf which describes the radar image noise r_{gL} becomes

$$r_{gL} = \frac{\sqrt{N} + 1.282}{\sqrt{N} - 1.282} \quad (6.40)$$

where:

N = number of looks averaged by the radar.

In our experiment we did not measure r_{gL} as a measure of the width of the noise pdf, rather we used the S/N ratio. Also, the two-dimensional spatial resolution was not measured directly rather the RMSB was obtained, however the spatial resolution can be approximated for a full focused system by

$$r_s = 1/\text{RMSB}. \quad (6.41)$$

Therefore, the SGL volume used in our experiment is defined as

$$V = \frac{1}{(\text{RMSB})(S/N)} \quad (6.42)$$

In addition a modified SGL volume was defined as

$$V_m = \frac{1}{(\text{RMSB})(S/N)(D)} \quad (6.43)$$

where:

D = Dynamic range.

The modified SGL volume attempts to account for the variation in dynamic range introduced into the degraded images. The original exponential model was also slightly modified to allow for a possible bias so the nonlinear model used here was

$$I = \alpha + e^{-\gamma V} \quad (6.44)$$

Because of the results of the correlation and linear regression analysis only response categories 1 and 5 will be reported. Table 14 contains the parameter estimates for categories 1 and 5 along with the SSE. From this data and plots of V_m and V versus the interpreter responses we concluded that this nonlinear model does not provide an acceptable description of the interpreter responses for this experiment.

Table 14

Image Quality Based on the Spatial-Grey-Level Volume

	Category			
	1		5	
	SGL Volume	Modified SGL Volume	SGL Volume	Modified SGL Volume
α	.765	2.4	3.17	2.61
β	3.75	147.0	39.7	4.0
γ	-.01	-.11	-.18	-.01
SSE	1004.4	1152.0	864.0	740.5
S^2	3.25	3.69	2.73	2.34

7.0 CONCLUSIONS

The purpose of this study was to investigate the relationship between measurable properties of radar images and the utility of those images for specific information extraction tasks. It was hoped that such a relationship would be useful for system design and image simulation. Five measurable image properties were identified, dynamic range, signal-to-noise ratio, image bandwidth, geometric fidelity, and root-mean square error. These metrics were determined to be linked to independent characteristics of the radar image and were either directly or indirectly related to many image quality parameters proposed in the past.

Clearly there are no physical laws governing the functional dependence of image utility upon these image metrics. therefore an experiment was conducted to empirically estimate a functional form for describing this dependence. This experiment consisted of obtaining SAR imagery and digitally processing it to create a set of radar images with controlled levels of image "quality". These images were presented to human interpreters who were asked to evaluate the usefulness of each image for extracting four classes or categories of terrain features, linear features, natural area features, complex area features, and individual man-made targets. In addition to these absolute responses or rankings of the images, the interpreters also rank ordered the images from best to worst relative to his evaluation of the utility of each image for assessing vehicle movement and activity level. Thus for each radar image the interpreter provided five responses. Also for each radar image each of the five image metrics were calculated. A functional relationship between each response category and the image metrics was

then estimated.

The first conclusion drawn from this experiment was that statistically significant regression equations could not be obtained to relate extracting natural area features, and individual man-made targets to the quality of radar images as judged by the interpreters. A possible explanation for this is that the interpreters did not use uniform criteria for these response categories owing to the complexity of these image features. However, for the simpler categories, i.e., linear features and relative ranking, statistically significant regression equations could be estimated. (Only second order interactions were considered in the regression equation.) Next it was found that for these categories the full second-order model was required to maintain the statistical significance of the regression equation. Another conclusion reached by this experiment was that different image metrics assume varying levels of importance as the response category was changed. For example, bandwidth and the signal-to-noise ratio were the most important metrics in estimating the ability of an interpreter to extract linear features from radar images. Dynamic range was predominant for estimating how an interpreter would relatively rank the radar images. This observation has important ramifications for the application of image quality metrics for multi-mission sensor design. That is, the system designer will have to trade-off system performance as a function of the application. A further conclusion obtained was that the nonlinear regression equations based upon the SGL volume were not statistically significant given these experimental data.

Recent advances in SAR systems, specifically the availability of

digital SAR images obtained from spaceborne platforms and digital image processing technology signify that more information extraction tasks will be performed by computer. Thus it is recommended that a study of this type be conducted to determine the relationship between measurable image properties which are related to system parameters and the success of automated information extraction algorithms for radar. For example, how is the probability of correct classification of targets related to measurable image properties. Using the computer to measure the utility of the radar images will remove the major disadvantage of the current study, i.e., relying on human judgments.

REFERENCES

- Barnard, Thomas W., "Image Evaluation by Means of Target Recognition," Photographic Science and Engineering, vol. 16, no. 2, March-April 1972, pp. 144-150.
- Biberman, L.M., "Fallacy and Fact of Sampled Imagery Displays," Human Factors, vol. 16, no. 3, 1974.
- Brock, G.C., "Discussion and Evaluation: Reflections on Thirty Years of Image Evaluation," Photographic Science and Engineering, vol. 11, no. 5, Sept.-Oct. 1967, pp. 356-362.
- Corbett, D.G., N.D. Diamantides and R.H. Kanse, "Measurements and Models Relating Physical Characteristics of Images to Target Detection," Behavioral Science Laboratory, Wright-Patterson Air Force Base, December 1964.
- Cox, D.R., Planning of Experiments, John Wiley and Sons, Inc., New York, 1958.
- Craig, D.W. and M.L. Hershberger, "Synthetic Aperture Radar Operator Tactical Target Acquisition Research," Proceedings of the Human Factors Society, 21st Annual Meeting, 1977.
- DiCaprio, Gabriel R. and James E. Wasielewski, "Radar, Image Processing, and Interpreter Performance," Photogrammetric Engineering and Remote Sensing, vol. 52, no. 8, August 1976, pp. 1043-1048.
- Draper, N.R. and H. Smith, Applied Regression Analysis, John Wiley and Sons, Inc., New York, 1966.
- Frost, V.S., J.A. Stiles, K.S. Shanmugam, J.C. Holtzman and S.A. Smith, "An Adaptive Filter for Smoothing Noisy Radar Images," Proceedings of the IEEE, vol. 69, no. 1, January 1981.
- Goodman, J.W., "Some Fundamental Properties of Speckle," J. Opt. Soc. Am., vol. 66, no. 11, November 1976, pp. 1145-1150.
- Granger, E.M. and K.N. Cupery, "An Optical Merit Function (SQF), Which Correlates with Subjective Image Judgments," Photographic Science and Engineering, vol. 16, no. 3, May-June 1972, pp. 221-230.
- Haralick, R.M., personal communications, 1978.
- Harger, Robert O., Synthetic Aperture Radar Systems: Theory and Design, Academic Press, New York, 1970.
- Harger, Robert O., "Synthetic Aperture Radar System Design for Random Field Classification," IEEE Trans. on Aerospace and Electronic Systems, vol. AES-9, no. 5, September 1973, pp. 732-740.

- Hunt, B.R., "Digital Image Processing," Proceedings of the IEEE, vol. 63, no. 4, 1975, pp. 693-708.
- Kondo, K., Y. Ichioka and T. Suzuki, "Image Restoration by Wiener Filtering in the Presence of Signal-Dependent Noise," Applied Optics, vol. 16, no. 9, September 1977, pp. 2554-2558.
- Kozma, Adam and Charles R. Christensen, "Effects of Speckle on Resolution," J. Opt. Soc. Am., vol. 66, no. 11, November 1976, pp. 1257-1260.
- Linfoot, E.H., Fourier Methods in Optical Image Evaluation, Focal Press, New York, 1960.
- Mendenhall, W., The Design and Analysis of Experiments, Duxbury Press, Belmont, California, 1968.
- Mitchel, Ralph H., "SAR Image Quality Analysis Model," Ph.D. Thesis, University of Michigan, 1974.
- Mitchell, Richard L., "Models of Extended Targets and Their Coherent Radar Images," Proceedings of the IEEE, vol. 62, no. 6, June 1974, pp. 754-758.
- Moore, R.K., "Trade-Off Between Picture Element Dimensions and Noncoherent Averaging in Side-Looking Airborne Radar," IEEE Trans. Aerospace Electronics, vol. AES-15, no. 5, September 1979.
- Myers, R.C., Response Surface Methodology, Allyn and Bacon, Inc., Boston, 1971.
- Naderi, F. and A.A. Sawchuk, "Detection of Low-Contrast Images in Film Grain Noise," Applied Optics, vol. 17, no. 18, 15 September 1978, pp. 2883-2891.
- Noffsinger, Edward B., "Image Evaluation: Criteria and Applications," S.P.I.E. Journal, vol. 9, November 1970, pp. 32-37.
- Noffsinger, Edward B., "Image Evaluation: Criteria and Applications, Paper 2: The MTF Criterion," S.P.I.E. Journal, vol. 9, March 1971, pp. 95-103.
- Oppenheim, A.V., R.W. Schafer and T.G. Stockham, Jr., "Nonlinear Filtering of Multiplied and Convolved Signals," Proceedings of the IEEE, vol. 56, August 1968, pp. 1264-1291.
- Penn, W.A., "Signal Fidelity in Radar Processing," IRE Transactions on Military Electronics, vol. MIL-6, April 1962, pp. 204-218.
- Porcello, Leonard J., Norman G. Massey, Richard B. Innes and James M. Marks, "Speckle Reduction in Synthetic Aperture Radars," J. Opt. Soc. Am., vol. 66, no. 11, November 1976, pp. 1305-1311.
- Pratt, W.K., Digital Image Processing, John Wiley and Sons, New York, 1978.

Soliday, S.M. and J.A. Gardner, "Picture Quality Judgments in a Digital Television System," Human Factors, vol. 16, no. 2, 1974.

Stockham, Thomas G., Jr., "Image Processing in the Context of a Visual Model," Proceedings of the IEEE, vol. 60, no. 7, July 1972, pp. 828-842.

Walkup, John F. and Robert C. Choens, "Image Processing in Signal Dependent Noise," Optical Engineering, vol. 13, no. 3, May-June 1974, pp. 258-266.

Wetz, J.M., "Criteria for Judging Adequacy of Estimation by an Approximating Response Function," Ph.D. Thesis, University of Wisconsin, 1964.

Williges, R.C. and C.W. Simon, "Applying Response Surface Methodology to Problems of Target Acquisition," Human Factors, vol. 13, no. 6, 1971.

Zelenka, Jerry S., "Comparison of Continuous and Discrete Mixed-Integrator Processors," J. Opt. Soc. Am., vol. 66, no. 11, November 1976, pp. 1295-1304.

APPENDIX A
INTERPRETER PACKET

INTRODUCTION

The purpose of this study is to establish quantitative techniques for predicting interpreter performance when an imaging radar is used as the reconnaissance sensor. This study consists of several phases. First, quantitative radar image quality factors were derived, then radar imagery was processed to exhibit controlled levels of image quality. Next, empirical data is to be gathered on interpreter performance vs. image quality. Last, this data will be used to derive the relationship between quantitatively measured image quality and interpreter performance.

Your contribution to this study is to interpret the given radar imagery following the instructions and guidelines and thus to provide the required empirical data. Also as part of the study the questions dealing with interpretation experience need to be answered.

The original radar imagery was collected by an X-band sensor with HH polarization and resolution of approximately 15 ft. with the look direction always left to right. There are four different target areas used in this study. The numbering scheme assigned 100-199 and 500-599 to Scene A, 200-299 and 600-699 to Scene B, 300-399 and 700-799 to Scene C, and 400-499 and 800-899 to Scene D. The target scenes are in the Roanoke, Virginia area.

Ancillary data are also provided in the form of aerial photographs (scale \approx 1:16000, taken in March 1968), USGS maps (scale 1:24000, 1968) and enlargements of the original SAR imagery (scale \approx 1:16000, collected in June 1968). Also provided will be lists of targets of interest contained within each scene and definition of the ranking criteria for each target.

You should use the ancillary data provided to locate each indicated target within each scene, then use the guidelines for ranking the quality of each target signature. We greatly appreciate your assistance in this study.

INSTRUCTIONS

- A. Interpret the images in the order presented.
- B. Use all available standard photo-interpretation equipment.
- C. Use the air photos, maps and radar images provided to locate each target within the processed radar images.
- D. For each scene rank only those targets which are indicated as being in that area.
- E. Provide the numerical ranking for each target signature according to the given definitions and guidelines.
- F. Be consistent in judging each image, that is, use the given guidelines for evaluating each target.
- G. Take at least a 10 minute rest after every hour of interpretation.

QUESTIONS

Please answer the following questions:

1. Number of years radar image interpretation experience? _____
2. General mission objectives of your image interpretation work?
3. List all photo-interpretation equipment used for the analysis of these test images.
4. Time used for this study? _____

GUIDELINES

These guidelines provide definitions for each target and its interpretation levels. In general the highest interpretation level of the six available is identification which indicates that the target can be distinguished from all others similar to it, e.g., an automobile bridge instead of just a bridge. This ranking is divided into two levels of certainty--possible and probable. A probable identification is when it is probable that the target signature can be correctly identified while a possible identification is when the target signature can possibly be correctly identified.

The next interpretation level is classification. Classification is being able to group a target return into a category, for example, being able to classify a signature as a bridge. Again this level is divided into two levels of certainty. A probable classification is when a signature can probably be grouped into a category while a possible classification is when a signature can possibly be correctly grouped into a category.

The lowest level is detection. Detection is when there is enough target signature to determine that there is something there while not being able to either identify or classify it. A possible detection is when it is possible that there is a target signature present while a probable detection is when it is probable that a target signature is present. No detection is when no target signature is present.

The procedure for attaching one of these interpretation levels to each target feature will be as follows:

1. Check to see if the target is present in the scene.
2. If it is present, next find its location in the ancillary photos, SAR images and maps and in the test image.

3. Next refer to the definition of the numerical rankings for this target.
4. Examine this target signature in the test image.
5. Assign a numerical ranking of identifiability based on the supplied definitions.

For example, consider the airport category. The ancillary data are used to find the location of its target signature within the degraded image. If its target signature is not present in the degraded image then a numerical ranking of 0 is assigned. A ranking of 1 is used if its target signature is possibly present while a ranking of 2 is assigned if its target signature is probably present. If the target signature can possibly be classified as an airport then a ranking of 3 is used, while if the signature can probably be classified as an airport then a ranking of 4 is assigned. If further information about the target signature can be obtained then a ranking of 5 or 6 will be used. Specifically, if the application of the airport can possibly be identified then a ranking of 5 is used, while if the application of the airport can probably be identified then a ranking of 6 is assigned.

TARGET	6 PROBABLE IDENTIFICATION	5 POSSIBLE IDENTIFICATION	4 PROBABLE CLASSIFICATION	3 POSSIBLE CLASSIFICATION	2 PROBABLE DETECTION	1 POSSIBLE DETECTION	0 NO DETECTION
All Roads	Target signature is probably a specific road configuration. e.g., Paved two lane or divided highway.	Target signature is possibly a specific road configuration.	Target signature is probably a road.	Target signature is possibly a forest/trees	There is probably a target signature.	There is possibly a target signature.	There is no target signature.
All Forest	Target signature is probably a specific forest class e.g., Deciduous trees.	Target signature is possibly a specific forest class	Target signature is probably forest/trees.	Target signature is possibly an agricultural area, but individual fields may not be discernible.	There is probably a target signature.	There is possibly a target signature.	There is no target signature.
Agriculture	Target signature can probably be divided into individual cultivated fields or clusters of fields.	Target signature can possibly be divided into individual cultivated fields or clusters of fields.	Target signature is probably a river/stream.	Target signature is possibly a river/stream.	There is probably a target signature.	There is possibly a target signature.	There is no target signature.
Rivers/streams	Stream/river order* probably can be specified in the image.	Stream/river can possibly be specified in this image.	Target signature is probably an airport.	Target signature is possibly an airport.	There is probably a target signature.	There is possibly a target signature.	There is no target signature.
Airport	Target signature is probably an airport used for a specific application, e.g., civilian or military.	Target signature is possibly an airport used for a specific application.	Target signature is probably a railway.	Target signature is possibly a railway.	There is probably a target signature.	There is possibly a target signature.	There is no target signature.
Railroad yard	Target signature can probably be separated into individual railway features, e.g., tracks, trains, roundhouse.	Target signature can possibly be separated into individual railway features.	Target signature is probably an industrial area.	Target signature is possibly an industrial area.	There is probably a target signature.	There is possibly a target signature.	There is no target signature.
Industrial Area	Target signature is probably a specific type industry, e.g., manufacturing, food processing, etc.	Target signature is possibly a specific type industry.					

* Stream order refers to a hierarchical description of the relationship of minor tributaries to major streams (or rivers). For example, fingertip tributaries along the watershed of a basin are first order streams; two first order s-reams converge to form a second order stream; two second order streams converge and form a third order stream, and so on.

TARGET	6 PROBABLE IDENTIFICATION	5 POSSIBLE IDENTIFICATION	4 PROBABLE CLASSIFICATION	3 POSSIBLE CLASSIFICATION	2 PROBABLE DETECTION	1 POSSIBLE DETECTION	0 NO DETECTION
Residential Areas	Target signature can probably be divided into individual dwell- ings of a resi- dential area.	Target signature can possibly be divided into indi- vidual dwellings of a residential area.	Target signature is probably a residential area.	Target signature is possibly a residential area.	There is probably a target signature.	There is possibly a target signa- ture.	There is no target signature.
Commercial Area	Target signature is probably a commercial area of a specific type, e.g., shopping center.	Target signature is possibly a commercial area of a specific type.	Target signature is probably a commercial area.	Target signature is possibly a commercial area.	There is probably a target signature.	There is possibly a target signa- ture.	There is no target signature.
Buildings	Target signature can be separated into a specific type of building, e.g., school, office building, apartment house, etc.	Target signature can possibly be separated into a specific type of building.	Target signature is probably a building.	Target signature is possibly a building.	There is probably a target signature.	There is possibly a target signa- ture.	There is no target signature.
Bridges	Target signature is probably a spe- cific type of bridge, e.g., rail or auto.	Target signature is possibly a specific type of bridge, e.g., rail or auto.	Target signature is probably a bridge.	Target signature is possibly a bridge.	There is probably a target signature.	There is possibly a target signa- ture.	There is no target signature.
Storage Tanks	Target signature is probably a spe- cific type of storage tank, e.g. water or oil.	Target signature is possibly a specific type of storage tank.	Target signature is probably a storage tank.	Target signature is possibly a storage tank.	There is probably a target signature.	There is possibly a target signa- ture.	There is no target signature.
Parked Vehicles	Target signature indicates specific type of vehicles such as automobi- les, airplanes, bus- es, tanks, etc.	Target signature is possibly a specific type of vehicle.	Target signature indicates parked vehicles of unde- termined type.	Target signature is possibly a parked vehicle.	There is probably a target signature.	There is possibly a target signa- ture.	There is no target signature.
Railyard Round House	Target signature indicates railroad roundhouse asso- ciated with loco- motive mainten- ance and repair.	Target signature is possibly a railroad round- house.	Target signature indicates presence of a large build- ing associated with a rail yard.	Target signature is possibly a roundhouse.	There is probably a target signature.	There is possibly a target signa- ture.	There is no target signature.

ANSWER SHEET

Interpreter Identification # _____
Image Identification # _____

Numerical Rankings

Interpretation Categories	Numerical Rankings								Target Location Number			
	6	5	4	3	2	1	0	A	B	C	D	
Linears Mostly Along Track^a												
Roads												
Paved												
Two Lane.....								•	•	•	•	
Four Lane.....									•			
Linears Mostly Cross Track^b												
Roads												
Paved												
Two Lane.....								•	•	•	•	
Four Lane ^d									•			
Linears Mostly Diagonal^c												
Roads												
Paved												
Two Lane.....								•	•	•	•	
Four Lane.....									•			
Unpaved												
Two Lane.....								•		•	•	
Divided Highway.....								•		•		
Natural Area Features												
Forest												
Deciduous.....									•			
Mixed.....										•	•	
Agriculture.....								•		•		
Rivers and Streams.....									•		•	
Complex Area Targets												
Airports.....										•		
Railroad Yards.....									•		•	
Industrial Areas.....									•		•	
Commercial Areas.....									•	•	•	
Residential Areas.....								•	•	•	•	
Individual Man-Made Targets												
Buildings.....								•	•	•	•	
Bridges.....									•		•	
Storage Tanks.....											•	
Parked Vehicles.....										•		
Railroad Round House.....											•	

- For these ratings indicate in the appropriate square whether the identification was based upon direct evidence or from the target context.
- ^a Angle of intersection of road and the flight path = 0-15°
- ^b Angle of intersection of road and the flight path = 75-90°
- ^c Angle of intersection of road and the flight path = 15-75°
- ^d Roads wide enough to carry four lanes of traffic are considered four lane even if the outside two lanes are used for parking

Interpreter Identification # _____

Separate the images by scene and rank them from best to worst in terms of the given applications.

For vehicle movement consider the potential for the movement of military vehicles, e.g. heavy trucks or tanks, through the given terrain. Terrain features which are important for assessing vehicle movement include bridge capacity, forest density, topographic slopes, water bodies and their depth, soils, street size, and road types, e.g. paved and unpaved. Evaluate and rank the images with respect to the quality of each image for specifying vehicle movement potential.

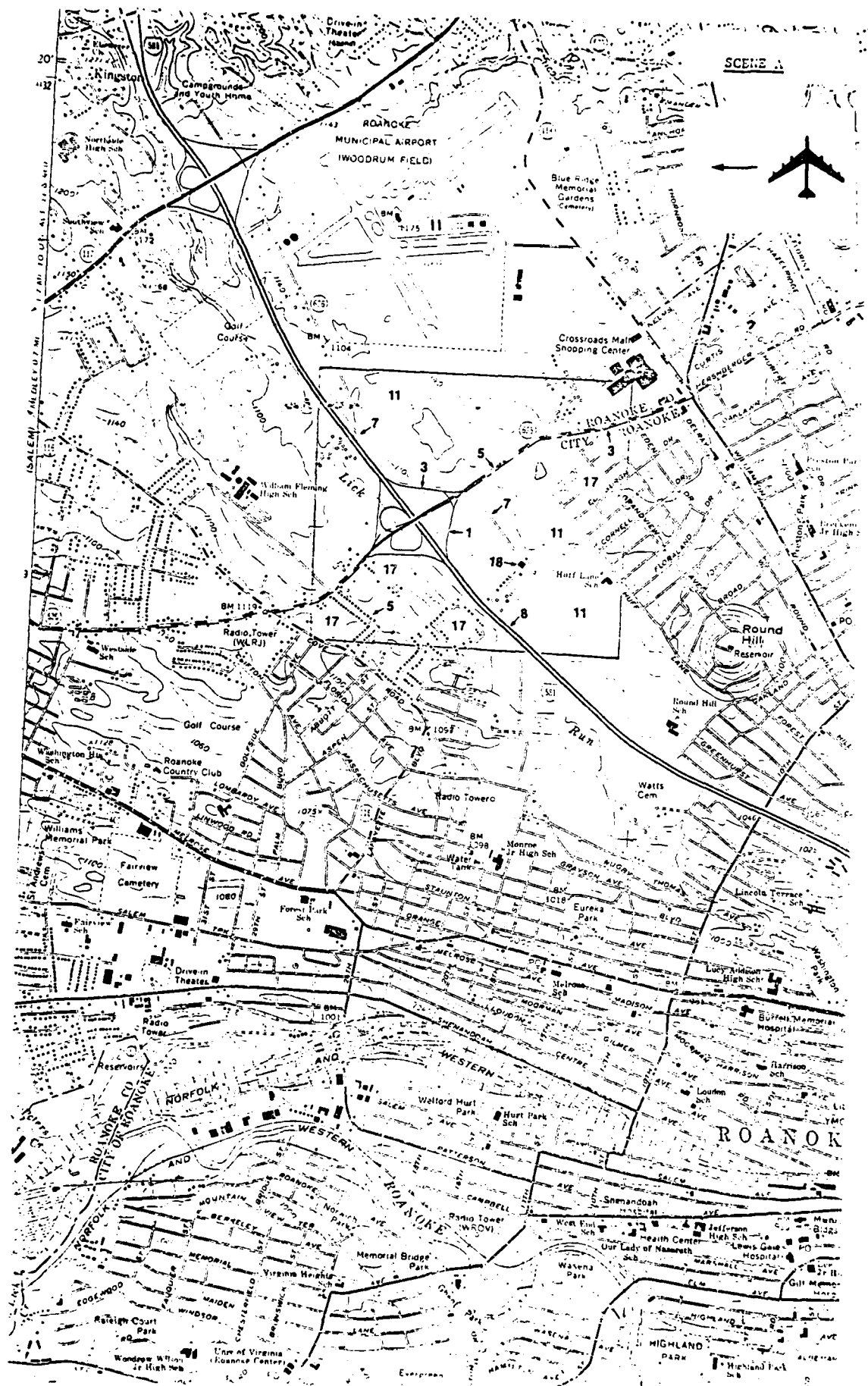
For activity level consider your ability to detect specific levels of activity. That is, order the degraded images from best to worst based on your assessment of the quality of each image for detecting levels of activity. For example, the relative quality of each image for detecting a convoy on the roads, a large increase or decrease in aircraft at the airport, or a large increase or decrease of rail cars at the railyard.

Scene A	Vehicle Movement	Activity Level
Best - 1		
Worst - 10		

Scene B	Vehicle Movement	Activity Level
Best - 1		
Worst - 8		

Scene C	Vehicle Movement	Activity Level
Best - 1		
Worst - 10		

Scene D	Vehicle Movement	Activity Level
Best - 1		
Worst - 12		

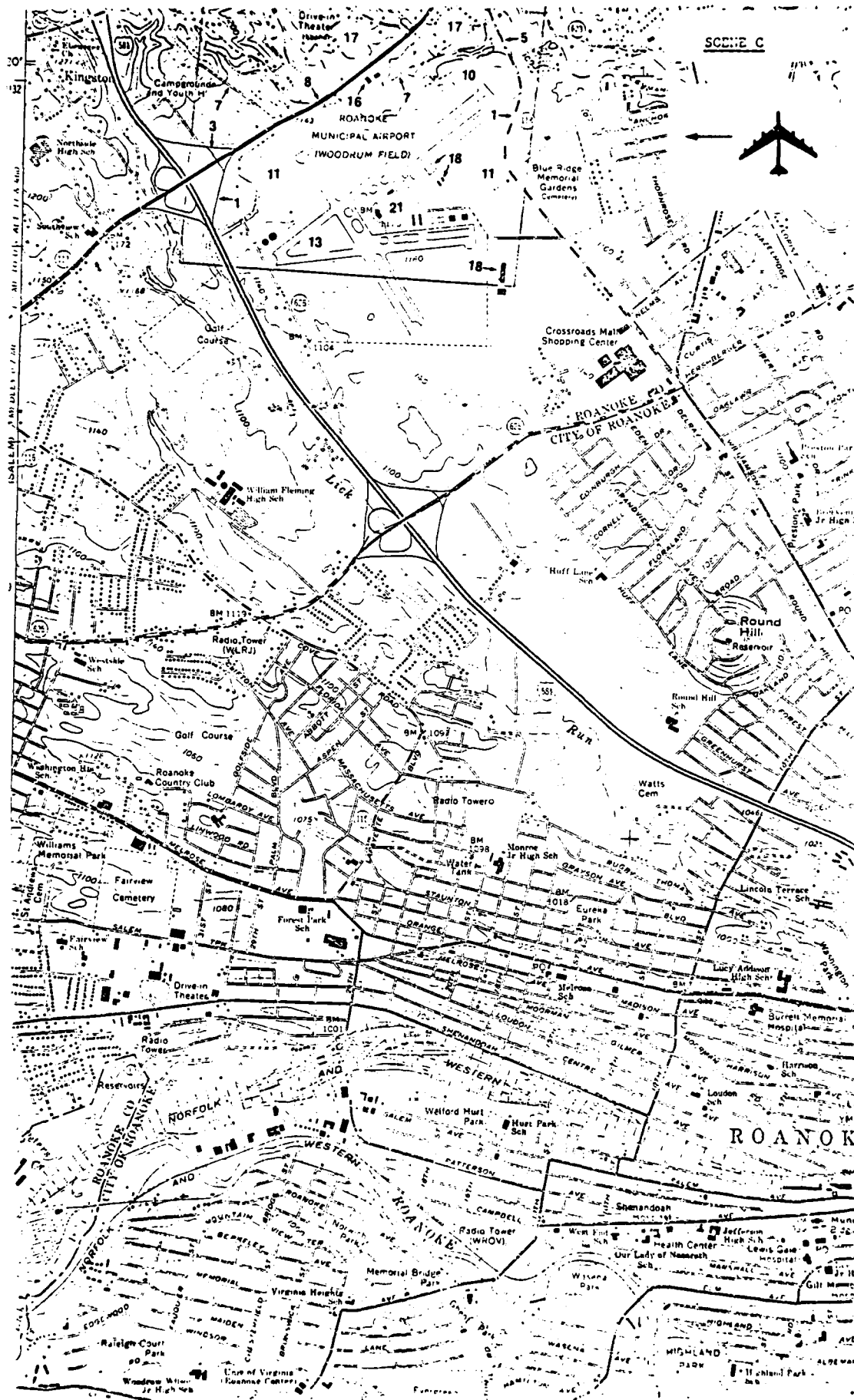


AD-A096 025 KANSAS UNIV/CENTER FOR RESEARCH INC LAWRENCE REMOTE --ETC F/G 17/9
RADAR IMAGE INTERPRETABILITY ANALYSIS.(U)
JAN 81 V S FROST, J A STILES, J C HOLTZMAN DAAG29-77-6-0075
UNCLASSIFIED CRINC/RSL-TR-342-2 ARO-14043.11-65 NL

2 of 2
NOB-120



END
DATE
FILMED
4-1-81
DTIC



APPENDIX B

PUBLISHED PAPERS AND SCIENTIFIC PERSONNEL
SUPPORTED BY RESEARCH AGREEMENT
DAAG29-77-G-0075

PUBLISHED PAPERS

- "Simulation of Imaging Radar Systems," (J. Holtzman, V. Kaupp, V. Frost, J. Abbott, and R. Martin), Eighth Annual Pittsburgh Conference on Modeling and Simulation, The University of Pittsburgh, Pittsburgh, Pennsylvania, April 21-22, 1977.
- "Image Synthesis for SAR Systems, Calibration, and Processor Design," (J.C. Holtzman, V.S. Frost, J.L. Abbott, and V. Kaupp), Proceedings of the Synthetic Aperture Radar Technology Convention, March 8-10, 1978, Las Cruces, New Mexico.
- "An Image Simulation Model for Radar Guidance," (J.C. Holtzman, V. Kaupp, V.S. Frost, J. Abbott), Ninth Annual Pittsburgh Conference on Modeling and Simulation, The University of Pittsburgh, Pittsburgh, Pennsylvania, April, 1978.
- "Radar Image Simulation," (J.C. Holtzman, V.S. Frost, J. Abbott, V. Kaupp), IEEE Transactions on Geoscience Electronics, Vol. GE-16, No. 5, October, 1978.
- "Development of Statistical Models for Radar Image Analysis and Simulation," (V.S. Frost), Master's Thesis, University of Kansas, 1978.
- "Computer Generated Radar Images for Navigation," (J.C. Holtzman, V.S. Frost, J.A. Stiles, and V. Kaupp), Proceedings of the Military Electronics Expo '78, Anaheim, California, November 14-16, 1978.
- "Seasonal Effects on Radar Imagery as Predicted by the PSM Simulation Techniques," (J.C. Holtzman, V.S. Frost, J.A. Stiles, E.E. Komp, E.S. Bergan, and V.H. Kaupp), Tenth Annual Pittsburgh Conference on Modeling and Simulation, The University of Pittsburgh, Pittsburgh, Pennsylvania, April, 1979.
- "Radar Image Simulation: A Project to Develop a Model, Define its Operational Constraints, Validate its Accuracy and to Produce Sample Results," (V.H. Kaupp), Doctor of Engineering Thesis, University of Kansas, May, 1979.
- "A Digital Computation Technique for Radar Scene Simulation: New SLAR," (J.C. Holtzman, V.S. Frost, J.L. Abbott, E.E. Komp, V.H. Kaupp), Simulation, June, 1979.
- "Radar Image Preprocessing," (J.C. Holtzman, V.S. Frost, J.A. Stiles, D.N. Held), Sixth Purdue Symposium on Machine Processing of Remote Sensing Data, Purdue University, West Lafayette, Indiana, June 2-6, 1980.
- "Digital Preprocessing of SEASAT Imagery," (J.C. Holtzman, J.A. Stiles, V.S. Frost, D.N. Held), International Conference on Communications, Seattle, Washington, June 8-11, 1980.

"Radar Image Enhancement and Simulation as an Aid to Interpretation and Training," (J.C. Holtzman, J.A. Stiles, V.S. Frost, L.F. Dellwig and D.N. Held), International Symposium on Remote Sensing of Environment, San Jose, Costa Rica, April 23-30, 1980.

<u>Scientific Personnel</u>	<u>Title</u>	<u>Degree Earned</u>
J.C. Holtzman	Principle Investigator	
V.H. Kaupp	Senior Project Engineer	Doctor of Engineering (Spring 1979)
V.S. Frost	Research Engineer	Master of Science (Fall 1978)
J.A. Stiles	Research Engineer	
S.A. Smith	Research Scientist	
C.J. Hahn	Research Technician	
D. Manweiler	Research Technician	

